REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
May 1995

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
Validity of Impairment and Functional Limitation Measures As Indicators and predictors of Disability following Acute Ankle Sprains Among Athletes

5. FUNDING NUMBERS

6. AUTHOR(S)
Rick W. Wilson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
AFIT Students Attending:
University of Virginia

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/CI/CIA
95-009 D

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DEPARTNEMT OF THE AIR FORCE
AFIT/CI
2950 P STREET, BDLG 125
WRIGHT-PATTERSON AFB OH 45433-7765

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release IAW AFR 190-1
Distribution Unlimited
BRIAN D. GAUTHIER, MSgt, USAF
Chief Administration

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

19950606 005

14. SUBJECT TERMS

15. NUMBER OF PAGES
150

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT

18. SECURITY CLASSIFICATION OF THIS PAGE

19. SECURITY CLASSIFICATION OF ABSTRACT

20. LIMITATION OF ABSTRACT

VALIDITY OF IMPAIRMENT AND FUNCTIONAL LIMITATION MEASURES AS INDICATORS AND PREDICTORS OF DISABILITY FOLLOWING ACUTE ANKLE SPRAINS AMONG ATHLETES

Rick W. Wilson
Charlottesville, Virginia

Bachelor of Science, University of Maryland, 1977
Master of Science, University of Alabama at Birmingham, 1980

A Dissertation Presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Doctor of Philosophy

Department of Education

University of Virginia

May 1995

ABSTRACT

This study considers the validity of measures commonly employed to assess disability and predict treatment outcome within sports rehabilitation research and practice. The concept of disability is developed using a model adopted by the Institute of Medicine of the National Academy of Science. Using this model, examples of treatment outcome measures taken from the sports medicine literature are classified into groups representing 4 domains of disablement: pathology, impairment, functional limitation, and disability. The relationships observed when this conceptual model is applied to examples chosen from the sports medicine literature support the theory that the effects of organic dysfunction on disability outcomes are mediated by behavioral factors. However, these observations are based on bivariate correlations between measures taken several years following knee injuries. Additional evidence concerning the construct, criterion-referenced, and evaluative validity of measures is required to determine whether this measurement theory may be more generally applied to sports rehabilitation practice and research.

To compare the usefulness of impairment and functional limitation measures in a sports medicine setting, 21

collegiate athletes were measured at 3 and 10 days post Grade I or II ankle sprain. Ankle pain, swelling, and range of motion were used as impairment indicators. Motor activity scores and a perceived athletic ability measure were used to indicate functional limitation. The number of days of athletic participation lost due to injury (mean = 11.9 $\pm$ 6.6 days) was used as the criterion measure of disability.

Evidence of construct validity obtained from hierarchical multiple regression and path analyses revealed that the functional limitation measures accounted for 98% of the explained variance (67% of the total variance) in disability duration. Impairment measures accounted for only 50% of the total variance. The functional limitation measures yielded more accurate predictions of disability duration, indicated by the standard error of the mean predicted values ($\pm$ 1.5 days), than organic measures ($\pm$ 2.0 days). The functional limitation measures also demonstrated patterns of time-dependent covariance indicating evaluative (longitudinal construct) validity and responsiveness to changes occurring between the 2 occasions of measurement.

These findings favor the use of behavioral measures of motor dysfunction over measures of organic dysfunction as indicators and predictors of disability following acute ankle sprains among athletes.

Department of Education

Graduate School of Arts and Sciences

University of Virginia

Charlottesville, Virginia

APPROVAL OF THE DISSERTATION

This dissertation, Validity of Impairment and Functional Limitation Measures as Indicators and Predictors of Disability Following Acute Ankle Sprains Among Athletes, has been approved by the Graduate Faculty of the University of Virginia in partial fulfillment of the requirements for the degree Doctor of Philosophy.

Joe H. Gieck, EdD, Advisor

Bruce M. Gansneder, PhD

Frank C. McCue, III, MD

David H. Perrin, PhD

Ethan N. Saliba, PhD

4/11/95 Date

DEDICATION

To Susan, Matt, Chris, and Amy, who gave patience and kindness when I insisted on having things my way.

> "Love bears all things, believes all things,
> hopes all things, endures all things." 1 Cor 13.7

ACKNOWLEDGEMENTS

No feat which we may accomplish and no goal which we may attain is ever achieved without the help of someone else along the way. I surely could not have completed this project without the generous gifts of time contributed by my committee members, Joe Gieck, Bruce Gansneder, Dave Perrin, Ethan Saliba, and Frank McCue. Your dedication and commitment to teaching and learning are deeply appreciated. I am also grateful for the enthusiasm and creativity shared by my doctoral student collaborators at the University of Virginia. Your company has been stimulating and inspirational, and I will miss you all. I take particular pleasure in acknowledging the assistance of my friend and respected colleague Dave Martin who, as site manager for the EPI ankle study, provided access to subjects and equipment during this project. Peace and long life.

/RWW May 95

TABLE OF CONTENTS

	Page
Chapter One, Introduction	
Statement of the Problem	1
Definition of Terms	4
Statement of Purpose	6
Research Questions	7
Delimitations	8
Limitations	8
Chapter Two, Review of the Literature	
Defining Functional Ability in Sports Medicine	9
Measuring Functional Ability in Sports Medicine	13
Impairment	13
Functional Limitation	17
Validity of Functional Ability Measures	28
Content Validity	28
Construct Validity	40
Criterion-referenced Validity	51
Usefulness of Functional Ability Measures	55
Predictive Measures	56
Discriminative (prescriptive) Measures	57
Evaluative Measures	61
Summary	66

Page

Chapter Three, Methodology

Subjects 74

Procedures 74

Impairment Measures 75

Functional Limitation Measures 77

Data Analyses and Calculations 80

Construct Validity 80

Predictive Validity 82

Longitudinal Construct (evaluative) Validity 83

Measurement Stability 84

Responsiveness 86

Chapter Four, Results

Reliability Study 88

Test-retest Reliability 88

Internal Consistency 88

Validity Study 90

Construct Validity 90

Predictive Validity 99

Responsiveness 99

Longitudinal Construct (evaluative) Validity 104

Page

Chapter Five, Discussion

Validity 108

Construct Validity 108

Predictive Validity 109

Reliability 112

Test-retest Reliability 112

Internal Consistency 115

Responsiveness 116

Conclusions and Recommendations 118

References 125

Appendix A 141

Appendix B 144

Appendix C 145

Appendix D 146

Appendix E 147

Appendix F 148

LIST OF TABLES

		Page
1.	Tegner Activity Score	20
2.	Noyes Sports Activity Rating Scale	23
3.	Marshall Scoring Scale	31
4.	Modified Larson Knee Scoring Index	33
5.	Iowa Athletic Knee Rating Scale	35
6.	Noyes Knee Rating Questionnaire	37
7.	Content of Lower Extremity Functional Rating Instruments Reported the Sports Medicine Literature	39
8.	Relationships Between Physical Impairment and Self-Report Measures of Functional Limitation Among Knee Injured Subjects	42
9.	Relationships Between Physical Impairment and Observed Functional Limitation Measures Among Knee Injured Subjects	45
10.	Relationships Between Observed and Self-Report Measures of Functional Limitation Among Knee Injured Subjects	48
11.	Relationships Between Clinical Indicators and Criterion Measures of Disability	53
12.	Intraclass Correlation Coefficients and Standard Errors of Measurement	89
13.	Internal Consistency of the Motor Activity Scale	91
14.	Hierarchical Regression of Disability Duration on Impairment and Functional Limitation Variables	94
15.	Accuracy of Mean Predicted Values of Disability Duration	100

Page

16. Responsiveness of Impairment and Functional Limitation Measures: Results of Paired t-tests 102

17. Interoccasion Stability of Mean Predicted values of Disability Duration: Results of Paired t-test 103

LIST OF FIGURES

	Page
1. Nagi's Model of the Disablement Process	12
2. Operational Model of the Disablement Process	27
3. Summary of Relationships (median values of r) Between Disablement Indicators Reported in the Sports Medicine Literature	69
4. Operational Model of Disablement Following Acute Ankle Sprain	92
5. Fully Recursive Model with Path Coefficients	96
6. Causal Model with Path Coefficients	98
7. Plot of Observed and Predicted Disability Duration (days lost) Showing 95% Confidence Intervals Around Mean Predicted Values	101
8. Longitudinal Construct Validity: Standardized Regression Coefficients Expressed as Percentages of Explained Variance in Disability Duration on 2 Occasions of Measurement	106

CHAPTER ONE

INTRODUCTION

Statement of the Problem

Measurement validity is a pivotal concept in both the research process and clinical practice of sports rehabilitation. Rehabilitation professionals employ a variety of clinical measures to assess the consequences of injury and determine an injured athlete's functional ability. However, there is confusion among researchers and clinicians regarding the definition of functional ability, and disagreement exists concerning which clinical measures are most useful in determining an athlete's rehabilitation status and forecasting readiness to return to competition following lower extremity injury.

A diverse assortment of functional ability measures have been proposed in the literature. However, sports medicine suffers from the absence of a unifying conceptual framework in which to evaluate and compare the validity of these measures. Researchers tend to emphasize precisely instrumented physical measures when assessing treatment effects, frequently dismissing self-reports and other behavioral measures as inherently unreliable. However, a

scientific approach to functional assessment should consider all appropriate dimensions of functional ability, including relevant behavioral and environmental factors, provided they can be quantified in a manner which produces trustworthy scores. Unfortunately, the scientific quality of potentially useful symptomatic and behavioral measures of disability have not been firmly established.

Investigators have also failed to relate scores obtained from either experimental or clinical measures to recognized criterion measures of clinical outcome. For example, even common physical measures of tissue and organ dysfunction, such as swelling or joint range of motion loss, have not been linked to disability duration following acute ankle sprain. These validity-related measurement deficiencies suggest that studies employing unproven and conveniently chosen dependent measures define functional ability too narrowly and may not accurately represent clinically relevant treatment effects.

Lacking proven measures for making accurate estimates of functional ability, rehabilitation professionals use empirical indicators, clinical "rules of thumb", to guide their decision making. When selecting measures to monitor rehabilitation progress or predict return to participation,

clinicians are forced to rely on "sensibility", or face validity (Feinstein, 1987). However, reliance on face validity alone to select a measure of functional ability invites disagreement and confusion. What is sensible to one clinician may be nonsense to another. Clearly, quantitative data demonstrating a measure's reliability, relationship to other measures, predictive accuracy, and sensitivity to changes in an athlete's functional status are preferable to face validity as criteria for selection of functional ability indicators. Unfortunately, the validity of clinical measurement systems intended to evaluate rehabilitation status, judge treatment effectiveness, and predict safe return to participation following lower extremity injury among athletes have not been established.

Definition of Terms

> "Fundamental to the development of theory - propositions about how important clinical concepts relate to one another and scientifically testing those propositions - is conceptual clarity. Only with conceptual clarity will we find accurate terminology and language with which to think, write, and speak about the concepts fundamental to the practice of physical therapy. Unfortunately, terminology used in the field of disablement research continues to breed confusion within and across disciplines." (Jette, 1994)

Assessment is the process of collecting data in order to make decisions. Assessment may include qualitative data.

Construct validity concerns the degree to which a characteristic relates to other variables as expected within a conceptual model (a set of theoretical relationships).

Content validity concerns the degree to which an instrument covers the extent of meanings within the concept it purports to measure and, simultaneously, the degree to which it is free of contamination by irrelevant factors.

Disability is the limitation or inability to perform a desired activity, task or role at an accustomed level in its normal context.

Evaluative validity refers to the parallel relationship between longitudinal changes in the indicator variables and the underlying trait of interest.

Face validity refers to that quality of an indicator which makes it seem a reasonable measure of some variable or attribute based on logic, clinical experience, or knowledge of biomedical principles.

Functional limitations are restrictions or lack of ability to perform a particular activity at the multisystem or total organism level.

Impairment refers to an anatomical abnormality or physiological dysfunction at the tissue or organ level.

Interrater reliability is the degree to which scores obtained by one trained observer agree with those obtained by other trained observers.

Longitudinal construct validity. (See evaluative validity)

Measurement is the assignment of numbers to a phenomenon or activity according to a set of rules.

Measurement theories consist of sets of rules for assigning numerals or magnitudes to an individual's characteristics or behaviors.

A test is a composite measure consisting of a set of items (activities or questions) presented to elicit a sample of performance on a single attribute or characteristic of interest. For example, the scores obtained from a motor performance sample are often used to estimate motor ability.

Statement of Purpose

The purpose of this investigation was to determine the validity of functional ability estimates generated by physical and behavioral measures taken from a sample of collegiate athletes with acute ankle sprains. The theoretical relationships between selected physical, self-report and motor activity measures specified by a proposed conceptual model of disability were tested for construct validity using data gathered in a clinical setting. The clinical usefulness of the conceptual model was evaluated on

the basis of criterion-referenced (predictive) validity and patterns of time-related covariance (responsiveness to change and longitudinal construct validity) observed among the measures.

Research Questions

The following research questions were proposed:

1. Does a conceptual model of disablement intended for a general population fit data gathered from athletes with acute grade I and II ankle sprains?

2. How accurate are predictions of disability duration based on motor behavioral measures compared to predictions based on physical measures following ankle inversion sprain among athletes?

3. Are selected measures of impairment and functional limitation responsive to change during the rehabilitation period following acute ankle inversion sprain?

4. Are the relationships observed among indicator variables consistent across occasions of measurement during the rehabilitation period following acute ankle sprains?

5. How reliable are measures used to indicate organic and motor behavioral dysfunction following acute ankle inversion sprains?

Delimitations

Participation in this study will be limited to college athletes who have sustained Grade I and II ankle inversion sprains less than 48 hours prior to their registration.

Limitations

The following limitations are acknowledged:

1. The results of this investigation will be characteristic of the participating subjects and may not be generalizable to other individuals or groups.

2. The results of this investigation may not be generalizable to other conditions or settings.

CHAPTER TWO

LITERATURE REVIEW

Defining Functional Ability in Sports Medicine

The concept of functional ability is essential to sports rehabilitation practice because athletes are not necessarily managed solely on the basis of medical diagnosis, but on the extent to which a condition prevents the athlete from performing. For a variety of reasons, including biological, behavioral, and environmental factors, individuals having the same diagnosis (e.g.- Grade II ankle inversion sprain or ACL deficient knee) may have differing abilities to function in athletics. The term "disablement" has been used to describe the various consequences of injury resulting in loss of function (Jette, 1994).

The use of the term disabled to describe the status of an injured athlete probably seems strange to most sports medicine clinicians. However, the concept of disability is inclusive of all socially defined roles and tasks, including:

> "Limitations in performance of such roles and tasks as related to family, work, community, school, recreation...etc." (Nagi, 1991, p.322)

Clearly, recreational and competitive sports require performance of specialized motor tasks, and sports participation may define distinctive roles for athletes, both on and off the playing field. According to this definition, a disabled athlete is one who is unable to participate in sports, or whose participation in sports is limited, due to pathology. The magnitude of the resulting limitation will be referred to here as disability or conversely, functional ability.

The inventory of measures commonly used to assess functional ability among athletes is both extensive and diverse. However, no universally recognized content standards for disability assessment instruments exist and new rating schemes appear in the literature each year. These instruments are published in a variety of journals over many years, making it difficult for clinicians to find and critically review their quality and relevance. Furthermore, there currently is no well-articulated conceptual framework in the sports medicine literature in which to compare measures of functional ability. However, two models, the World Health Organization's (WHO) International

Classification of Impairments, Disabilities, and Handicaps (ICIDH) (WHO, 1980), and the Nagi classification model (Nagi, 1991) are frequently referenced in the general rehabilitation literature (Heerkens, 1994; Jette, 1994; Duncan, 1994; Pope & Tarlov, 1991). Both of these conceptual schemes use four distinct but interrelated domains to describe disablement. The ICIDH uses four classifications: disease, impairment, disability, and handicap. In the Nagi model (Figure 1), the four categories are termed pathology, impairment, functional limitation, and disability.

Within Nagi's model, disablement is viewed as a process which flows unidirectionally across these four domains from pathology to disability. However, depending on the circumstances, the degree of functional limitation or disability experienced by the injured athlete is not necessarily proportional to the severity of the pathology. Efforts of the sports medicine staff, distinctive characteristics of the athlete, and environmental factors can all act to modify disability at any of the preceding stages (Pope & Tarlov, 1991). Compared to the WHO model, Nagi's taxonomy offers some advantages in terms of its conceptual clarity, and it has subsequently been approved

Pathology --->	Impairments --->	Functional Limitations --->	Disability
(diagnoses of disease, injury, or condition)	(dysfunctions and structural abnormalities in specific body systems)	(restrictions in physical and mental performance at the level of the whole organism)	(limitation or inability to perform normal roles, tasks and activities to levels expected within physical and social contexts)

Figure 1. Nagi's Conceptual Model of the Disablement Process

and published with slight modification by the National Academy of Science's Institute of Medicine(Pope & Tarlov, 1991). This conceptual framework will be used to define functional ability and describe the disablement process within the confines of this study. In order to better understand the measurement of functional ability, particularly as it pertains to athletes with lower extremity injuries, a more detailed discussion of the components of Nagi's disablement model is needed.

Measuring Functional Ability in Sports Medicine

Impairment

According to Nagi's classifications, impairment addresses anatomic abnormality and tissue, organ or system dysfunction in response to injury or disease. Indicators of impairment are found in the attributes of the individual (Nagi, 1991) and can be grouped into two broad categories: physical signs and self-reported measures (symptoms).

Physical Signs of Impairment. Physical impairment measures are derived from examination or observation, and examples include instrumented and subjective evaluations of joint stability, arthroscopic examination of tissue integrity, electromagnetic imaging, passive joint range of

motion, and other anthropometric data such as girth measurements or volumetric displacement.

Goniometric measurements of joint range of motion are often employed in the clinical assessment of impairment. Stability of these measures at the ankle was reported by Elveru and colleagues (1988) who examined the reliability of passive ankle and foot goniometric measures using intraclass correlation coefficients (ICCs) derived from measurements obtained from 43 orthopedic patients. Intratester ICCs (1,1) (Shrout & Fleiss, 1979) were 0.90 for dorsiflexion and 0.86 for plantar flexion. However, the objectivity of ankle range of motion measurements is moderate at best. Intertester ICCs (1,1) were 0.50 for dorsiflexion and 0.72 for plantar flexion. Data regarding the precision of the goniometric measures obtained during this study were not reported. In explaining their results, the authors commented that goniometric measurement may be useful when a single clinician is evaluating a patient by comparing one limb with the other or evaluating changes in a patient's range of motion over time (Elveru et al., 1988).

Volumetric measurement of limb swelling is based on the ancient observations of Archimedes, who noted that the volume of water displaced by a submerged object corresponds to the volume of the object submerged. This method of

physical assessment is used extensively in clinical and research settings to document swelling, particularly of the hands and feet. The precision of foot volumetry as a clinical measure of impairment is well documented. Goldie and colleagues (1974) reported foot volumetry yielded a coefficient of variation [CV = (SD/Mean) x 100] of 0.4%. Smyth and associates (1963) found a similar degree of precision in a series of 5 foot volume measurements taken from 4 individuals in rapid succession on the same day. The 8 CVs from this study (one from each limb) averaged 0.35%, with no single intrasubject measurement error exceeding 0.6%. The same investigators also reported a high degree of stability for volumetric measurements, observing daily foot volume fluctuations in normal adults of less than 1.5% over a 5 day period (Smyth et al., 1963).

Self-report Measures of Impairment. An athlete's inability to perform may be related to any number of symptoms, such as joint pain or "giving way", which may accompany injury or disease. Indeed, the inability to perform desired physical activities without pain or instability is a major reason why patients seek medical advice (Hutchinson et al., 1979), and restoration of normal athletic activity without symptom imposed limitations is the

goal of successful treatment and rehabilitation in sports medicine. As Giove and colleagues (1983) point out, not only the ability to perform, but also the perceived quality of that performance should be considered in disability assessment (my emphasis added).

> "The ability to be active *in a symptom-free manner* is the goal of most patients. Thus, 'success' to the patient is a return to those activities which are important to him or her. Perhaps more cogent is the fact that "failure" to the patient is the inability to return to certain activities." (Giove et al., 1983)

Patients' perceptions of joint pain, swelling, and instability are included in several questionnaires designed to measure disability following lower extremity injury (Jensen et al, 1983; Marshall et al, 1977; Noyes et al., 1989; Tegner & Lysholm, 1985; Walla et al, 1985). Unfortunately, a thorough review of the sports medicine literature indicates that the reliability of symptom scores produced by these instruments has not been reported.

Functional Limitation

In Nagi's taxonomy, functional limitation refers to the limitation or inability to perform an activity. Although both impairment and functional limitation involve function, in functional limitation the frame of reference is at the

level of the person as a whole rather than at the tissue or organ level. Personality traits and other attributes of the individual may intervene to modify the effect of pathology and impairment on functional limitation measures. Candidate indicators of functional limitation following lower extremity injury include locomotor activities such as walking, running, squatting, stair climbing or jumping (Pope & Tarlov, 1991). As with impairment, functional limitation measures can be divided into 2 categories, observed and self-reported, for the purpose of discussion.

Observational measures of functional limitation. A diverse array of motor activities have been proposed as indicators of functional ability following lower extremity injury. Kettlekamp and Thompson (1975) scored running, jumping, and cutting activities. Arnold and associates (1979) described the "cross-over" test as a method of detecting rotational instability of the knee. Daniel and colleagues (1982) included a shuttle run, while Curl and associates (1982) added a circle run and vertical jump. The Tegner performance rating system incorporated a figure-8 run, stair running, and a downhill slope run (Tegner et al., 1986). Tibone and Antich (1988) proposed use of the "straight cut" and "cross cut" maneuvers in addition to the

full squat, half squat, and running in place. The "hop-test", introduced to the literature by Daniel and colleagues (1988), was later augmented and evaluated by Noyes' Cincinnati group (Barber et al., 1990; Noyes et al., 1991). Lephart and colleagues (1992) advocated the use of a three item "functional performance test" consisting of shuttle run, carioca, and dynamic quadriceps/hamstring cocontraction activities to assess functional ability following ACL injury.

Noyes and colleagues (1980) recognized closed kinetic chain performance measures as indicators of functional ability, noting that these motor activities incorporate components of both static and dynamic joint stability. Isokinetic torque production and other open kinetic chain "strength" measures are also appropriately included in this classification because motor performance can be modified by both reflex mechanisms and the amount of voluntary inhibition or effort exercised by the subject (Lankhorst et al., 1985).

Self-report measures of functional limitation. Feinstein (1987) observed that medicine, perhaps more than any other discipline, uses scales to measure attributes such as health status or disability. The unique characteristic of

this type of instrument is that a single score indicates the athlete's status with regard to every classification (Babbie, 1990). For example, in a true scale of functional ability, component items are graded from "easy" to "hard". If the functional ability of interest were running endurance, three test activities might be "run five minutes on the track", "run ten minutes on the track", and "run 15 minutes on the track". The assumption is made that if an individual can perform a particular task, he or she could also perform all the easier subordinate tasks. Several researchers have developed activity rating scales to estimate functional limitation based on the severity of symptoms reported by patients in response to various hypothetical physical tasks.

Lysholm and associates (Lysholm & Gillquist, 1982; Odensten et al., 1983a; Tegner & Lysholm 1985) developed an activity rating questionnaire (Table 1) which included a hierarchy of work and sport activities. Reasoning that differences in functional ability can depend on the activity level at which the symptoms are regarded as "significant" by the patient, the questionnaire was designed to determine the self-reported activity threshold at which a patient's knee became symptomatic.

Table 1

Tegner Activity Score

10. Competitive sports Soccer-national and international elite 9. Competitive sports Soccer, lower divisions Ice hockey Wrestling Gymnastics 8. Competitive sports Bandy Squash or badminton Athletics (jumping, etc.) Downhill skiing 7. Competitive sports Tennis Athletics (running) Motorcross, speedway Handball Basketball Recreational sports Soccer Bandy and ice hockey Squash Athletics (jumping) Cross-country track findings both recreational and competitive 6. Recreational sports Tennis and badminton Handball Basketball Downhill skiing Jogging, at least five times per week	5. Work Heavy labor (e.g.,building, forestry) Competitive sports Cycling Cross-country skiing Recreational sports Jogging on uneven ground at least twice weekly 4. Work Moderately heavy labor (e.g., truck driving, heavy domestic work) Recreational sports Cycling Cross-country skiing Jogging on even ground at least twice weekly 3. Work Light labor (e.g., nursing) Competitive and recreational sports Swimming Walking in forest possible 2. Work Light labor Walking on uneven ground possible but impossible to walk in forest 1. Work Sedentary work Walking on even ground possible 0. Sick leave or disability pension because of knee problems

(Adapted from Tegner & Lysholm, 1985)

Flandry and colleagues (1991) also published a questionnaire which relied solely on self-reports to establish functional ability ratings. Using a system of visual analog scales (VAS), scores were recorded for 28 unweighted items related to both severity and frequency of pain, swelling, and giving way as well as self-reported difficulties accomplishing specific motor activities. In discussing potential advantages of visual analog response scales relative to ordinal scales, Flandry and colleagues (1991) stated that:

> "Having a response that ranges from 0 to 100 and is carried to as many decimals as measurement techniques allow boosts statistical power tremendously over question forms where only several categorical answers are available." (Flandry et al., 1991)

This comment probably refers to the increase in statistical power which accompanies the use of parametric rather than non-parametric techniques of data analysis. The decision to employ parametric statistics depends upon whether equal intervals exist between adjacent categories of possible numerical ratings on VAS response scales. This subject remains controversial. Feinstein (1987) generally classified analog scales as quasi-interval levels of measurement, suggesting that the burden of proof regarding the existence of interval level data in any given application ultimately

rests with the investigator. In this regard, Maxwell (1978) has validated the use of parametric statistics for VAS data analysis, while Price and colleagues (1982) demonstrated the ratio scaling properties of the VAS in studies of pain measurement. Furthermore, Monte Carlo style studies by Boneau (1960) have shown t-tests (and by inference, F-tests) to be remarkably robust to skewness, heterogeneity of variance and other violations of statistical assumptions so long as sample sizes are equal and share roughly the same distribution.

Noyes and colleagues (1989) published a Sports Activity Rating Scale to compare the results of treatment interventions between groups of individuals with knee problems. The logic employed in the construction of this instrument is that patients are assigned an activity score based on their frequency of athletic participation. Position within the frequency-defined region is then further determined by the type of activity engaged in by the patient. The Noyes Sports Activity Rating Scale (Table 2), is a four tiered hierarchical system based on frequency of participation in athletics (Noyes et al., 1989). Each of the four frequency levels are further divided into three additional levels defined by intensity of activity. A numerical value ranging from 0 to 100 is assigned to each of

Table 2

Noyes Sports Activity Rating Scale

	Points	Sports
Level I		
4-7 days/week	100	Jumping, hard pivoting, cutting (basketball, volleyball, football. soccer, gymnastics)
	95	Running. twisting. turning (racquet sports, baseball, hockey, skiing, wrestling)
	90	No running, twisting. jumping (running. cycling, swimming)
Level II		
1-3 days/week	85	Jumping, hard pivoting, cutting (basketball, volleyball, football, soccer, gymnastics)
	80	Running, twisting. turning (racquet sports, baseball, hockey, skiing, wrestling)
	75	No running, twisting, jumping (running, cycling. swimming)
Level III		
1-3 times/month	65	Jumping, hard pivoting, cutting (basketball, volleyball, football, soccer. gymnastics)
	60	Running, twisting, turning racquet sports. baseball, hockey, skiing, wrestling)
	55	No running. twisting, jumping (running. cycling.swimming)
Level IV		
No sports possible	40	ADL with no problems
	20	ADL with moderate problems
	0	ADL with severe problems

(Adapted from Noyes et al., 1989)

the twelve resultant categories. However, to produce reliable scores, a scale must contain a response set in which classifications are both discrete and mutually exclusive (Babbie, 1980). Noyes' practice of simultaneously grading functional limitation by two criteria, frequency and type of activity, violates these measurement principles and leads to scores which appear to lack sensibility. Based on Noyes' classification scheme, an athlete who engages in swimming four times a week receives a score of 90/100, while one who plays football or basketball 3 times a week receives only 85/100 points. Does swimming four times per week actually represent a higher functional activity level than playing basketball three days per week? No statistical evidence has been found to indicate that the scores obtained from the Noyes Athletic Rating Scale are either reliable or valid indicators of functional ability.

The three self-report measures of functional limitation identified here are derived from multi-item questionnaires containing an apparent hierarchy of physical activities. However, a hierarchical structure of responses based on face validity does not insure that a true scale of functional limitation exists. Face validity refers to that quality of an indicator which makes it seem a reasonable measure of some variable or attribute based on logic, clinical

experience, or knowledge of biomedical principles (Feinstein, 1987; Johnston et al., 1992).

Other criteria are preferable to face validity when evaluating the quality of rating scales. Internal consistency is essential to the reliability of multi-item instruments when the items are added up or labeled as assessing the same underlying attribute, but this property must be demonstrated through statistical analysis of the patterns of actual item responses (Babbie, 1990; Hulin et al., 1983). The concept of internal consistency is concerned with the degree to which individual items within an instrument are related to the total score. Unless the relationship between individual items and the total score is demonstrated, the sum of unrelated items may not predict anything well, even if the individual items are correlated with functional ability. Consequently, the stability of the aggregate score may obscure variations within potentially important traits (Johnston et al., 1992). The data analysis techniques used to establish internal consistency of a scale include Cronbach's alpha, factor analysis techniques, and others. However, statistical evidence confirming the property of unidimensionality has not been discovered for any of the multi-item functional rating instruments presented in this review. Furthermore, while these

statistics can identify relationships between scale items, studies of internal consistency and structure are not adequate substitutes for studies of reliability concerning stability over time or agreement between raters (Johnston et al., 1992).

Functional limitation may be measured using either observational or self-report techniques. Observed and self-reported activity levels and behaviors may differ for a variety of reasons. The quality of data produced by self-reports depends to some extent on recall, interpretation, or conjecture on the part of the respondent or rater. Observational measures can be distorted by bias introduced by an observer's subjective interpretation of patient responses. Cook and Campbell (1979, p.66) point out that when all measures are presented in the same way or use the same means of recording responses, the method itself introduces a form of bias to a study. It would seem advantageous to include and compare data gathered by both methods in order to determine which indicators should be included in a comprehensive evaluation instrument intended to obtain an accurate and well-differentiated picture of functional ability following acute ankle sprain among athletes.

Pathology ----> Impairments ----> Functional Limitations----> Disability

	- Observed	- Observed
	- Self-reported	- Self-reported

Figure 2. Operational Model of Disablement Following Acute Ankle Sprain

Validity of Functional Ability Measures

Validity is the chief criterion for the appropriate selection and use of a measure. In the realm of measurement theory, validity is concerned with the degree to which an instrument or test measures the concept it was intended to measure. The definition of validity may appear to be simple, but it is probably the most difficult measurement concept to prove (Babbie, 1990).

In this section, three interrelated types of validity will be used to critique instruments identified in the sports medicine literature as candidate measures of functional ability following lower extremity injury. These three types, content, criterion-referenced, and construct validity, are not the only classifications of validity, "but have stood the test of time in the biometric and psychometric literature" as criteria for instrument selection (Johnston et al., 1992).

Content Validity

Content validity is the degree to which an instrument covers the extent of meanings within the concept it purports to measure, and examination of content is a primary means of selecting a measure for clinical use. Recall that Nagi's theoretical model defines disability in terms of four

domains. Functional ability can be estimated using either a single indicator or a combination of measures. However, it is unlikely that a single indicator can be identified which adequately describes the concept of functional ability. Because functional ability is a multidimensional concept encompassing biological, behavioral, and environmental variables, any single measure probably both underrepresents the total concept of functional ability and contains some degree of irrelevance to the same concept (Cook & Campbell, 1979). Several multi-item instruments found in the sports medicine literature include a combination of physical measures, observational measures, and symptom ratings to assess functional ability following lower extremity injury.

Content Validity of Composite Measures. To demonstrate content validity, an instrument should include measures which take into account all important domains or sources of variance which provide useful estimates of functional ability (Cook & Campbell, 1979; Kirshner & Guyatt, 1985). Early composite instruments designed to assess disability after lower extremity injury grew from attempts to compare patient outcomes following various treatment protocols. Specifically prominent were the efforts of orthopedic surgeons to identify knee injured patients who would be most

likely to benefit from surgical intervention based on a multi-item measures which have been generically termed "knee scores" (Marshall et al., 1977; Jensen et al., 1983; Lysholm & Gillquist, 1982; Noyes et al., 1984, 1989; Walla et al., 1985). Component items of the knee scores typically included physiological measures of organic impairment, observed motor activities, and patients' self-reported symptoms and functional abilities.

Marshall, Fetto, and Botero (1977) stated that the purpose for developing their composite assessment instrument was to establish a method of patient evaluation and follow-up after knee ligament injury which would allow the clinician to: (1) measure a patient's anatomical defects and functional deficits throughout the course of diagnosis and treatment, (2) assess the effects of various treatment modalities in a specific case, and (3) determine the eventual "functional toll", of a specific type of injury. These stated goals suggest that the authors were interested in developing a composite tool which would simultaneously assess physical impairment, functional limitations, and disability. The Marshall knee index (Table 3) consists of individual items scored on an ordinal scale, with the maximum number of points awarded to individual items ranging

Table 3

Marshall Scoring Scale

Pain	O = Yes 1 = No	Thigh sizes	O = >2 cm difference 1 = 1-2 cm difference 2 = Equal
Swelling	O = Yes 1 = No		
Stair difficulty	O = Yes 1 = No		
Clicking/numbness	O = Yes 1 = No		
Giving way	O = Regularly upon daily activities 1 = With stress upon daily activities 2 = With stress only 4 = Normal, none	Range of motion	O = <90° 1 = Limited flexion and extension 2 = Limited flexion or extension 3 = Normal
		Stability	
Return to sports/work	O = No return 1 = Return to different 2 = Return to original with limitations 3 = Full return	LCL	O = Gross instability 2 = Instability in flexion and extension 3 = Moderate instability in flexion 4 = Mild instability in flexion 5 = Normal
Functional tests			
Duck walk	O = Cannot perform 1 = Can perform but with discomfort 2 = Can perform		
Run in place	O = Cannot 1 = Can	MCL	O = Gross instability 2 = Instability in flexion and extension 3 = Moderate instability in flexion 4 = Mild instability in flexion 5 = Normal
Jump on one leg	O = Cannot perform 1 = Can perform but with discomfort 2 = Can perform		
Half squat	O = Cannot 1 = Can		
Full squat	O = Cannot 1 = Can	ACL	O = Severe in neutral and rotation (Pivot shift, Slocum, Jerk test) 2 = Severe in neutral 3 = Moderate jog 4 = Slight jog 5 = Normal
Specific knee examinations			
Tenderness	O = Yes 1 = No		
Joint effusion	O = Yes 1 = No		
Swelling	O = Yes 1 = No		
Crepitations	O = Yes 1 = No		
Muscle power	O = Very weak 1 = Diminished flexion and extension 2 = Diminished flexion or extension 3 = Normal	PCL	O = Severe in neutral and rotation 2 = Severe in neutral 3 = Moderate jog 4 = Slight jog 5 = Normal

Adapted from Tegner & Lysholm, 1985

from 1 to 5. Output from the Marshall scale is expressed as a total "knee score", obtained by summing point values from the individual items. Marshall and colleagues (1977) originally published their 50 point index along with outcome data obtained from 280 patients with knee ligament injuries.

Another reason to construct a multivariate clinical index of function is to develop an efficient and comprehensive means of collecting and organizing data (Babbie, 1990). Jensen and colleagues (1983) had this goal in mind when they developed a composite instrument to evaluate the effect of different variables on clinical outcome and determine the correlation between clinical tests and functional deficits. This instrument has subsequently been referred to as the Larson index (Table 4). The Larson index is a 200 point, weighted-item score introduced in a pre- and postoperative study of 205 patients who had surgical reconstruction of the anterior cruciate ligament (ACL) (Jensen et al., 1983). The 200 points were divided evenly between the "objective" (physical measures) and "subjective" (self-reported symptoms and activity levels) portions.

Another index, the Iowa Athletic Knee Rating Score (IAKRS) is a 100 point, global knee rating instrument (Wallaet al., 1985). Like the Marshall and Larson indices,

Table 4

The Modified Larson Knee Scoring Index

Function	(55 points)	Impairment	(45 points)
Gait (10 points)		**Pain** (30 points)	
Limp:		None	30
None	5	Not incapaciting	25
Slight	3	Incapaciting	10
Marked	0	Severe	0
Support:			
Full support	5	Swelling (3 points)	
Stick or crutch	3	None	3
Weight bearing impossible	0	Slight and occasional	1
		Frequent	0
Activities (45 points)			
Stairs:		Atrophy of thigh (2 points)	
No difficulty	10	1-2 cm	2
Slight difficulty	6	None	1
One step at a time	2	More than 2 cm	0
Unable	0		
Squatting:		Range of motion (10 points)	
No difficulty	5	0-45°, deduct for each	
Slight difficulty	4	10° loss	1
Not past 90°	2	45-90°, deduct for each	
Unable	0	15° loss	1
Walking:		90-100°, deduct for each	
Unlimited	20	20° loss below 130	1
More than 2 km	15		
1-2 km	5		
U nable	0		
Running:			
No difficulty	5		
Slight difficulty	4		
Straight ahead only	2		
Unable	0		
Jumping:			
No difficulty	5		
Slight difficulty	3		
Unable	1		

Adapted from Lysholm & Gillquist, 1982

the IAKRS combines patients' symptoms and activity grading with measured performance on motor tasks and clinical indicators of strength and joint stability into a summative knee score (Table 5). This instrument, like all of the "knee scores "mentioned so far, must be completed by a clinician, and 40% of its content consists of the clinician's interpretation of a physical examination. The IAKRS also contains an activity-referenced symptom index. This twelve item questionnaire asks subjects to rate and report their symptoms of pain, swelling, and giving way during four levels of physical activity: walking, running, mild twisting/pivoting, and hard twisting/pivoting. Another feature, the athletic activity rating portion of the index, is a 10 point hierarchical scale stratified by type of sport and level of competition.

Noyes and colleagues (Noyes and McGinniss, 1985; Noyes et al., 1984, 1989, 1990) developed and refined a functional rating questionnaire which includes self-report measures of both impairment and functional limitation. This instrument also differs from previous rating forms insofar as it is designed to be completed by the patient rather than a clinician. Functional ability is graded on a 100 point system, with 50 points awarded based on the intensity of pain and giving way reported during sports activity,

Table 5

Iowa Athletic Knee Rating Scale

		Total Points
Symptoms		30
Pain	10	
Swelling	10	
Instability	10	
Activity status		30
Athletic	10	
Work	20	
Physical examination		40
Laxity	21	
Range of motion	2	
Synovitis	3	
Crepitus	3	
Hamstring control	11	
		100

(Adapted from Walla et al., 1985)

recreational activities, or activities of daily living. The remaining 50 points are based on self-reported difficulty encountered during walking, running, stair climbing, and jumping/twisting. The most recent iteration of Noyes' Knee Function Rating Form (Table 6) also includes an overall knee condition rating.

All of the rating instruments discussed to this point were developed to gather and evaluate data following knee injury. Comparable literature pertaining to functional ability following ankle injury is considerably less extensive. Hocutt and associates (1982) recorded the number of post-injury days elapsed before subjects could perform four closed kinetic chain activities without "significant" pain or discomfort. These activities were: 1) standing, 2) walking, 3) stair climbing, and 4) running and jumping. However, these data were not analyzed to determine whether the functional levels described comprised a scale of unique and mutually exclusive functional states. Linde and colleagues (1984) developed a 48 point ankle rating index which included four functional levels: 1) inability to stand on the injured extremity, 2) inability to work, 3) inability to participate in sports, and 4) no impairment. This instrument also included measures of pain, swelling, and

Table 6

Noyes Knee Rating Questionnaire

PLEASE CHECK THE STATEMENT THAT BEST DESCRIBES THE CONDITION OF YOUR KNEE.

PAIN

20	I experience no pain in my knee.
16	I have occasional pain with strenuous sports or heavy work. I don't think that my knee is entirely normal. Limitations are mild and tolerable.
12	There is occasional pain in my knee with light recreational sports or moderate work.
8	I have pain brought on by sports, light recreational activities, or moderate work. Occasional pain is brought on by daily activities such as standing or kneeling.
4	The pain I have in my knee is a significant problem with activities as simple as walking. The pain is relieved by rest. I can't participate in sports.
0	I have pain in my knee at all times, even during walking, standing, or light work.

SWELLING

10	I experience no swelling in my knees.
8	I have occasional swelling in my knee with strenuous sports or heavy work.
6	There is occasional swelling with light recreational activities or moderate work.
4	Swelling limits my participation in sports and moderate work. Occurs infrequently with simple walking or light work. Occasionally with simple walking or light work--about three times a year.
2	My knee swells after simple walking activities and light work. The swelling is relieved by rest.
0	I have severe swelling with simple walking activities. The swelling is not relieved bv rest.

STABILITY

20	My knee does not give out.
16	My knee gives out only with strenuous sports or heavy work.
12	My knee gives out occasionally with light recreational activities or moderate work; it limits my vigorous activities, sports, or heavy labor.
8	Because my knee gives out, it limits all sports and moderate work. It occasionally gives out with walking or light work.
4	My knee gives out frequently with simple activities such as walking. I must guard my knee at all times.
0	I have severe problems with my knee giving out. I can't turn or twist without my knee giving out.

OVERALL ACTIVITY LEVEL

20	No limitations. I have a normal knee, and I am able to do everything including strenuous sports and/or heavy labor.
16	I can partake in sports including strenuous ones but at a lower level. I must guard my knee and limit the amount of heavy labor or sports.
12	Light recreational activities are possible with RARE symptoms. I am limited to light work.
8	No sports or recreational activities are possible. Walking activities are possible with RARE symptoms. I am limited to light work.
4	Walking activities and daily living cause moderate problems and persistent symptoms.
0	Walking and other daily activities cause severe problems.

WALKING

10	Normal, unlimited.
8	Slight, mild problems.
6	Moderate problem, flat surface up half a mile.
4	Severe problems, only 2-3 blocks.
2	Severe problems, need cane or crutches.

RUNNING

0	Normal, unlimited, fully competitive.
8	Slight, mild problems, run at half speed.
6	Moderate problems, only 1 -2 miles possible.
4	Severe problems, only 1-3 blocks possible.
2	Severe problems, only a few steps.

STAIRS

5	Normal, unlimited.
4	Slight mild problems.
3	Moderate problems, only 10-15 steps possible.
2	Severe problems, require banister for support.
1	Severe problems, only 1-5 steps with support.

JUMPING AND TWISTING

5	Normal, unlimited, fully competitive.
4	Slight, mild problems, some guarding.
3	Moderate problems, gave up strenuous sports.
2	Severe problems, affects all sports, always guarding.
1	Severe problems, onlv light activity possible (golf/swim).

If I had to give my knee a grade from 1 to 100, with 100 being the best, I would give my knee a ________

(Adapted from Noyes et al., 1989).

range of motion in the total score. However, neither validity nor reliability data have been found which would provide a basis for evaluating the scientific quality of this rating index. Table 7 summarizes the content of the composite functional rating instruments presented in the sports medicine literature and compares them to the four components of Nagi's conceptual model of disability: pathology, impairment, functional limitation, and disability.

Internal structure of composite measures. Each of the composite measures of functional ability found in the sports medicine literature derives a total score (e.g.- knee score) based on item summation. Each individual measure, activity, or question is given an item value which is then added to values for all the other items. This practice assumes that the characteristic of unidimensionality exists among the individual items. That is, all items must measure the same trait or characteristic if summation is to yield a meaningful score of that trait or characteristic. A valid assessment of a multidimensional concept such as functional ability should be comprehensive enough to address all of its important aspects. This suggests that a measurement instrument designed to assess functional ability may

Table 7

Content of Lower Extremity Functional Rating Instruments Reported in the Sports Medicine Literature

Index Name	Pathology Knee	Pathology Ankle	Physical Impairment	Functional Limitation	Disability	Source
Functional Performance Test	+			+	+	Lephart et al., 1992
Hughston Subjective Index	+		+	+		Flandry et al., 1991
Iowa Athletic Knee Rating	+		+	+		Walla et al., 1985
Larson Index	+		+	+		Jensen et al., 1983
Lysholm Score	+		+	+		Lysholm & Gillquist, 1983
Marshall Score	+		+	+	+	Marshall et al., 1977
Noyes Score	+		+	+		Noyes & McGinnis, 1985
Noyes Sports Activity Rating	+			+		Noyes et al., 1989
Tegner Activity Rating	+			+		Tegner & Lysholm, 1985
Linde Ankle Score		+	+		+	Linde et al., 1984
Hocutt Ankle Scale		+		+		Hocutt et al., 1982

contain several measures which may not be additive. However, the relationships between individual measures used to provide a comprehensive assessment of functional ability following lower extremity injury among athletes remain unclear.

Construct Validity

The degree to which a characteristic relates to other variables as expected within a conceptual model (a set of theoretical relationships) is construct validity (Babbie, 1990). Because constructs are phenomena with multiple attributes, evaluating construct validity involves piecing together a network of relationships between measures of those attributes (Johnston et al., 1992).

> "Specifically, construct validity is tested by (1) seeing whether a measure displays the pattern of converging or predictive relationships it should (convergent validity); (2) distinguishing the construct from confounding factors (discriminant or divergent validity); and (3) measuring with variations in settings, populations, and even details in measurement procedure so that generalization can be made beyond a narrow application. These requirements are not easily met." (Johnston et al., 1992)

Multivariate statistical procedures (e.g.- LISREL, factor analysis, MANOVA, path analysis) can be used to determine whether variables in a study or subdivisions of a composite

instrument relate to each other as expected or explained by a conceptual model.

Relationships between measures of impairment and functional limitation. Table 8 summarizes concurrent relationships between measures of physical impairment and perceived functional ability reported among knee injured subjects. In general, physical measures demonstrate a poor correlation with perceived functional limitation in these studies.

Lephart and colleagues (1992) found arthrometry scores (KT1000 @ 90N) to be poorly correlated ($r = 0.14$) with functional ability (IAKRS) scores among 41 ACL deficient athletes. Similar findings had been previously reported by Harter and associates (1988), who found subjective functional ability ratings (Noyes knee rating scores) which were independent ($r = -0.02$) of instrumented arthrometry (KT 1000 @ 90N) measures among 51 ACL reconstructed subjects. These findings also support the earlier work of Seto and colleagues (1988) who stated that 8 manual tests of static ligamentous stability demonstrated low to moderate correlations with functional activity scores following ACL reconstruction. Functional activity scores also failed to

Table 8

Relationships Between Physical Impairment and Self-Report Measures of Functional Limitation Among Knee Injured Subjects

Study	Impairment Measure	Functional Measure	Relationship
Harter	KT 1000	Noyes Score	r = -0.02
Seto	Knee ROM Stability Effusion	Functional Activity Score	No significant correlations (r values not stated)
Lephart	KT 1000	IAKRS	r = 0.14
Walla	ACL Laxity (pivot shift)	IAKRS	No significant relationship (r values not stated)

display significant, but unreported, correlation with knee effusion (Seto et al., 1988), thigh girth (Lephart et al.,1992), or knee range of motion (Seto et al., 1988; Lephart et al., 1992) in these studies.

Walla and colleagues (1985) employed the IAKRS in a study which also was designed to identify factors which were associated with the ability to return to vigorous physical activity without surgical intervention. Among 38 volunteer former athletes with histories of unreconstructed ACL injuries which occurred at least 24 months prior to the study, the total IAKRS score was moderately correlated to the patient's own estimates of functional ability ($r = 0.659$). The authors stated that no relationship was found between the functional rating score and either of two physical measures: thigh circumference or degree of ligamentous instability. Again, however, specific correlation coefficients were not presented with these data.

The findings that measures of biological dysfunction at the tissue level do not correlate well with perceived measures of functional limitation are somewhat surprising. These measures may represent independent attributes of disability. However, it seems unlikely that severity of injury would not be related to functional limitation. An explanation may be found in the dependent variables used in

these studies. All of the studies cited above rely on subjects' self-reported activity levels to define functional limitation. For a variety of reasons, self-reported and actual levels of athletic participation may differ, leading to potential measurement inaccuracies. Jette (1994) and Rothstein (1994) eloquently argue that observable measures should be used in studies designed to assess functional ability. Table 9 summarizes the relationships between impairment and observed measures of functional limitation.

Andersson and colleagues (1991) reported a moderate correlation ($r= -0.37$) between ligamentous laxity ratios, measured by instrumented arthrometry at 90 N, with observed functional ability (Tegner scores) in their study of 156 ACL deficient and ACL reconstructed subjects. Similar but somewhat lower associations were observed between knee arthrometry measurements and performance times obtained from a shuttle run ($r= -0.23$), cocontraction drill ($r= -0.21$), and carioca run ($r= -0.27$) among subjects with chronic ACL deficiencies (Lephart et al., 1992). Barber and colleagues (1990) also reported that no statistically significant relationships existed between KT-1000 arthrometer scores and performance results from five closed chain motor activities (single leg hop for distance, single leg vertical jump, single leg timed hop, shuttle run without pivot, and shuttle

Table 9

Relationships Between Physical Impairment and Observed Functional Limitation Measures Among Knee Injured Subjects

Study	Impairment Measure	Functional Measure	Relationship
Andersson	KT 1000	Tegner Score	r=-0.37
Barber	KT 1000	Hop Tests Shuttle Run	"No significant relationship"
Lephart	KT 1000	Shuttle Run Cocontraction Hop Test	r = -0.23 r = -0.21 r = -0.27
Odensten	ACL Laxity (pivot-shift)	Hop Test Figure 8 Run	No difference between stable/unstable groups
Risberg	KT 1000	Stair Hop Triple Hop	"correlated" (r value not given)

run with pivot) obtained from 35 ACL deficient subjects. Unfortunately, the authors of this study did not present the correlation coefficients obtained between these measures. The presence of a positive pivot-shift among 60 patients who had undergone distal iliotibial band transfers for ACL insufficiency failed to adversely affect performance on single leg hop tests or a figure 8 run (Odensten et al., 1983b). No relationship was found between motor performance and clinical examination findings or arthroscopy among 26 ACL deficient male subjects (Tegner et al., 1986).

In a more recent study, Risberg and Ekeland (1994) reported the degree of knee instability, indicated by KT-1000 arthrometry, and thigh girth measures both correlated with performance on single leg cross-over hop and stair hop tests among 35 post ACL reconstruction patients. The authors claimed that lower correlations were found between these physical measures and closed chain activities involving the use of both lower extremities (figure 8 run and stair run), suggesting that these bilateral activities may be appropriate indicators of ability to resume daily activities while the single limb tests may indicate ability to return to athletic activity. Again, however, these authors failed to publish the actual correlation coefficients obtained. Further, no data were presented which would support the

authors' premise that the ability to perform single limb activities is related to an athlete's demonstrated ability to return to participation.

Relationships between observed and self-reported functional limitation. Table 10 summarizes the degree of association reported between self-report and observed activity measures in the sports medicine literature. Andersson and colleagues (1991) found a moderate correlation (r=0.42) between self-reported physical activity ratings (Lysholm scores) and motor performance scores (Tegner activity score) among 156 ACL deficient and ACL reconstructed subjects. A similar relationship (r=-0.32) was observed between turn times obtained during a figure-8 run and Lysholm activity rating scores in this study.

The relationships between isokinetic torque production and self-reported functional ability among knee injured patients are generally characterized as weak to moderate. Regrettably, some investigators have chosen not to report the magnitudes of "nonsignificant" correlations between measures. Keller and associates (1993) reported "no relationship" (r values not stated) between isokinetic scores and self-reported functional ability among 40 patients with posterior cruciate deficient knees without

Table 10

Relationships Between Observed and Self-Report Measures of Functional Limitation Among Knee Injured Subjects

Study	Observed Measure	Self-report Measure	Relationship
Keller	Cybex II	Noyes Score	"Nonsignificant" (r values not reported)
Barber	Cybex Hop Tests	Instability during sports	"significant" (r values not reported)
Harter	Cybex II	Noyes Score	Q PT r = 0.24 HS PT r = 0.11
Wilk	Biodex Hop Tests	Noyes Score	Q PT r = 0.44 - 0.71 HS PT r = 0.18 - 0.39 r = 0.31- 0.51
Andersson	Figure 8 (turn time)	Lysholm Score	r = 0.42
Odensten	Cybex II	Lysholm Score	r = 0.47
Risberg & Ekeland	Stairs Run Figure 8 Run	Lysholm Score	Loading factor: -0.62
	Stairs Hop Triple Hop	Lysholm Score	Loading factor: -0.43
Lephart	Cybex II	IAKRS	Q PT r = 0.13 - 0.15 HS PT r = 0.17 - 0.19

presenting the supporting data. Barber and colleagues (1990)appear to contradict these findings by reporting "significant" correlations (r values not stated) between abnormal hop test performance, isokinetic knee extensor weakness and subjective instability during sports. Fortunately, other researchers have been more forthcoming.

Wilk and colleagues (1994) studied the relationships between isokinetic torque production and Noyes activity rating scores among 50 ACL reconstructed patients. Correlations between peak isokinetic quadriceps torque and Noyes scores were moderate, and the magnitude of the correlation coefficients was observed to decrease as the testing velocity was successively increased from 180 to 300 and 450 degrees/second. Interestingly, the opposite trend was observed during hamstring testing. Correlations between Noyes scores and peak hamstring torque values increased with increasing test velocity. All of the correlations between peak hamstring torque and self-reported functional limitations ($r= 0.18-0.39$) were lower than any of those reported for the quadriceps ($r= 0.44-0.71$). The tendency for Noyes scores to correlate more closely with quadriceps than hamstring torque among ACL reconstructed subjects had previously been reported by Harter and associates (1988). Furthermore, Noyes and associates (1991) had previously

observed that correlations between isokinetic quadriceps torque and hop test scores, like subjective ratings, tended to be greater at low velocities (60 deg/sec) than at higher ones (300 deg/sec) following ACL rupture.

Relationships between isokinetic torque and functional limitation, as measured by other self-report instruments, vary somewhat from those obtained using Noyes scores. Correlations between IAKRS scores and peak isokinetic torque of the quadriceps and hamstrings have been reported as "low" ($r = 0.13$-0.19) among ACL deficient subjects, but no discernible trend accompanied velocity changes between 60 and 270 degrees/second (Lephart et al., 1992). Odensten and colleagues (1983b) reported a correlation coefficient of 0.47 between Lysholm scores and peak isokinetic quadriceps torque at 30 degrees/second among 60 subjects following extra-articular stabilization surgery for ACL insufficiency.

Another observable measure of functional ability involves a group of closed chain activities called "hop tests". Wilk and colleagues (1994) reported moderate associations ($r = 0.38$-0.51) between hop test performance and self-reported functional assessment scores among 50 ACL reconstructed patients. In a similar study, Risberg and Ekeland (1994), observed that self-reported functional scores (Lysholm scores) were more closely related to shuttle

run performance than to triple jump or stair hop performance among 35 post ACL reconstruction patients.

Construct validity, even with the help of a theoretical model, is not easily proven (Campbell & Fiske, 1959; Johnston et al., 1992). The relationships between clinical indicators of impairment and functional limitation presented so far have been derived solely from bivariate correlations. When two measures fail to concur, it is unclear which one may depart from the conceptual model of disability. Although a high degree of association between two variables implies interchangeability, it does not indicate what is being measured. The major problem with these studies is the absence of a "gold standard" or criterion measure of functional ability.

Criterion-referenced Validity

Criterion-referenced validity refers to the degree to which a measure relates to, or predicts, some other characteristic or event. There are two types of criterion-referenced validity, concurrent and predictive, which are distinguished by the temporal relationships between observations of the independent and dependent variables. The first, concurrent validity, involves comparison of a measurement with supporting evidence which is gathered at

approximately the same time as the measure being validated. Predictive validity examines the justification of using a measurement to say something about future events or conditions.

Candidate indicators of functional ability demonstrate criterion-referenced validity if, and only if, they correlate with accepted measures of functional ability (Lankhorst et al., 1985). Fortunately, sports rehabilitation is a practical science with a generally accepted and readily observable functional outcome criterion. That criterion is safe return to athletic participation. However, the evidence correlating return to sports participation with impairment or functional limitation measures is meager and largely inconclusive. A thorough examination of the sports medicine literature produced only four studies which used return to athletic activity as the criterion measure of functional ability (Table 11).

Giove and colleagues (1983) measured athletic activity on a 4 tiered scale of self-reported sports participation. The instrument used to measure activity in this study combined features of the Noyes (Noyes et al., 1983) and Lysholm (Lysholm & Gillquist, 1982) subjective knee rating questionnaires. These investigators reported that ACL deficient individuals with isokinetic hamstring/quadriceps

Table 11

Relationships Between Clinical Indicators and Criterion Measures of Disability

Study	Indicator	Disability Criterion	Relationship
Parolie	KT 1000 (PCL)	Return to Sports (yes/no)	No significant relationship
Giove	HS/Quad Ratio	Return to Sports (4 levels of activity)	HS/Q ratio>1 scored "significantly" higher
Seto	Functional Activity Score		r = 0.78
Lephart	IAKRS Score Shuttle Run Cocontraction Hop Test	Return to Sports (yes/no)	"Significant" differences between groups (return vs. no return)

ratios greater than 1:1 reported significantly higher levels of sports participation than subjects exhibiting less hamstring torque relative to the quadriceps. Using a similar sports participation scale as the criterion measure, Seto and associates (1988) reported a positive correlation ($r=0.78$) between four self-reported levels of sports participation and activity from a sample of 24 ACL reconstructed subjects at an average postoperative interval of 5 years.

Lephart and colleagues (1992) used multiple measures to assess functional outcome in a study of 41 physically active patients with ACL deficient knees. The disability criterion in this study was return to sports participation 10-36 months post-injury. Independent variables included instrumented knee arthrometry (KT-1000), isokinetic tests of the quadriceps and hamstrings (Cybex), and three closed kinetic chain activities (shuttle run, carioca, and cocontraction tasks). A composite knee score (IAKRS) and motor performance measures appeared to best represent an individuals ability to return to athletic participation. Lephart and colleagues (1992) failed to observe significant correlations between instrumented measures of ligamentous laxity (KT 1000 @ 90N) and self-reported ability to resume accustomed athletic. Similar findings were reported by

Parolie and Bergfeld (1986), who also found no significant relationship (correlation coefficients were not reported) between instrumented arthrometry (KT 1000 @ 90N) scores and self-reported return to sports participation in a study of 25 posterior cruciate ligament deficient athletes.

Usefulness of Functional Ability Measures

As previously noted, validity is the chief criterion for selecting a measure of functional ability. However, the validity of a measure depends not only on the scientific quality of the instruments under consideration, but also upon the purpose of the measurement. In this sense, we may only refer to a measurement as being valid for a particular purpose. Furthermore, the relevant criteria for validity vary with the purpose and application of a particular measure (Rothstein & Echternach, 1993).

Generally speaking, functional ability may be assessed for three purposes: (1) to predict the duration or severity of disability; (2) to classify individuals according to severity or type of functional disabilities for treatment purposes, or; (3) to evaluate change in disability over time or in response to treatment. Measures designed for one of these purposes can be problematic if used for a different

one. Those who wish to use a particular instrument to gather data for clinical decision making or research should understand how certain characteristics of measures can affect their suitability for the intended purpose (Kirshner & Guyatt, 1985).

Predictive Measures

A predictive measure is one which is used to forecast an event, such as return to participation, or estimate an individual's status with regard to some criterion measure (Kirshner & Guyatt, 1985; Law, 1993). For example, an athlete's readiness to return to practice is often predicted or estimated based on physical measures or the ability to perform particular motor activities. Although any variable may be used for prediction, not all measures are equally well-suited to the purpose. The predictive validity of a measure is related to its ability to estimate values of the dependent variable more accurately and consistently than predictions arrived at by chance (Kirshner & Guyatt, 1985).

To demonstrate predictive validity, correlation coefficients are used to assess the relationship between measures taken on one predictor variable and a second dependent variable. When more than one predictor variable is used, regression analysis is commonly employed to determine

the relative weights of variables used to predict treatment outcome or estimate the magnitude of the dependent variable.

Discriminative (prescriptive) Measures

Sometimes, the clinician/researcher needs to distinguish levels of functional status between individuals or groups when no criterion measure or gold standard is available. For instance, the investigator may need to categorize research subjects according to functional ability for between-groups comparison, while the clinician may use a set of classification criteria to decide whether an athlete would benefit from a particular rehabilitation procedure. To assist with these decisions, it would be advantageous to select a measure which is sensitive to differences between individuals (Kirshner & Guyatt, 1985). To be considered valid, the classifications produced by a discriminative measure must demonstrate the variance present between subjects or groups of subjects in the sample. A reliable discriminative measure of functional ability must produce between-subjects variance which is stable between measurements, provided actual functional status has not changed. Therefore, a valid and reliable discriminative index must produce between subjects variance which is large

and which is also stable between measurements (Kirshner & Guyatt, 1985).

For example, Tegner and colleagues (1986) reported that 26 ACL deficient male subjects single-leg hopped significantly shorter distances than a comparison group of 66 male soccer players. Time taken to negotiate the turn portion of a figure-8 course, slope running, and stair running were also significantly greater in the patient group. These activities appear to be good candidates as items in a discriminative index of functional ability because they are sensitive to differences between injured and normal athletes. However, most of the injured athletes performed within normal range on straight running and one-leg hop, and 35-47% of them scored normally on turn running, stairs, and slope running. These findings suggest inadequate discriminative power of the former two tests, making them less useful to clinicians and researchers interested in identifying functional limitations among athletes.

In an effort to establish prescriptive validity for the single hop and timed hop tests, data from 35 ACL deficient subjects published by Barber and colleagues (1990) and another sample of 67 ACL deficient subjects were examined by Noyes and associates (1991). Using a criterion of less than 85% normal symmetry between knees to identify ACL deficient

subjects yielded a false positive prediction rate for ACL deficiency of 3%-6% (PPV = 0.86-0.92, specificity = 0.97) and a false negative prediction rate of 48% (NPV = 0.74-0.84, sensitivity = 52%) for single hop performance. Data from the timed hop tests were subjected to the same analysis with the following results: false positive prediction rate for ACL deficiency was 6% (PPV = 0.71-0.85, specificity 94%), while false negative prediction rates were 51-57% (NPV = 0.72-0.82, sensitivity = 43-49%). Due to the relatively high rates of false negative predictions and low sensitivity values computed for these tests, the authors recommended that they not be used as the sole determinants of ACL insufficiency during clinical examination. The discriminant validity of these motor ability scores with regard to athletic performance was not examined in this study.

Lephart and colleagues (1992) did compare the ability of multiple measures to discriminate between categories of athletic performance in their study of 41 physically active patients with ACL deficient knees. The disability criterion in this study was return to sports participation (yes/no) at 10-36 months post-injury. Discriminative variables used in this study included instrumented knee arthrometry (KT-1000), isokinetic tests of the quadriceps and hamstrings (Cybex), and three closed kinetic chain activities (shuttle run,

carioca, and cocontraction tasks). A composite knee score (IAKRS) which contained physical measures, athlete's perceptions of functional ability, and motor performance measures appeared to best represent an individual's ability to return to athletic participation. Based on these findings, the authors proposed that a combination of motor activities and an athlete's perceptions of functional ability be used to decide whether a return to participation is appropriate. However, they offered no guidance concerning how indicators of motor performance or perceived functional ability should be weighted or summarized to accurately determine whether a particular individual could be expected return to participation.

With the permission of the author, Lephart's data (1991) were reanalyzed using SPSS^x DISCRIMINANT to estimate the usefulness of closed kinetic chain performance scores for this purpose. Results indicated an accurate assignment to self-reported "return" or "non-return" groups for 78% of the cases. A significant chi squared test (df=1, p=.01) led to a decision to reject the null hypothesis that these classifications were no better than those arrived at by random assignment to the groups (prior probability =.50 per group). Additional computations revealed 33% false positive (PPV = 0.90, specificity = 67%) and 19% false negative (NPV

= 0.67, sensitivity = 81%) classification rates. These discriminant analysis statistics, combined with those of Barber and associates (1989), suggest that closed chain motor activities may be more closely correlated to performance deficits (disability) than to anatomical deficits (impairment) among ACL deficient subjects.

Evaluative Measures

According to Kirshner and Guyatt (1985), an evaluative instrument is one which measures change in a variable within a patient or group over time. Measures of this type are potentially useful indicators of rehabilitation progress. The usefulness of evaluative measures is improved if they are relatively unaffected by measurement error, are inexpensive enough to be used frequently and are resistant to reactivity, particularly test-retest learning or fatigue effects (Johnston et al., 1991).

Lysholm and Gillquist (1982) had this purpose in mind when they developed the Lysholm activity rating score.

> "The main advantage of the (Lysholm) activity rating scale is not to compare different patients, but to note changes in activity level in the same person at different times." (Tegner et al., 1988)

The Lysholm score measures patient progress by establishing a symptom related activity score before treatment and

comparing it with the score attained at the end of rehabilitation. This process results in a raw score which is a simple index of change (Feinstein, 1987). The Lysholm score is based on the absolute ability to either perform or not perform certain physical tasks. In a recent study, Risberg and Ekeland (1994) evaluated the relationship between Lysholm activity rating scores and 6 closed chain performance tasks (vertical jump, figure-8, stairs running, triple jump, stairs hop, and side jump) among 35 post-ACL reconstruction patients (X=18 months post-op). Factor analysis appeared to establish single-leg and bilateral closed chain activities as independent measures of functional ability. Lysholm activity rating scores correlated with bilateral closed chain activity performance. The authors speculate that Lysholm scores may demonstrate a "ceiling effect". That is, the Lysholm score may demonstrate the intraindividual variability in activity level found relatively early in rehabilitation, but may not be sensitive to higher functional levels represented by more vigorous activities encountered later on. However, this study is typical of previous investigations of functional ability in that it relies on a single measurement occasion or pretest/posttest design to produce data. These designs do not allow adequate opportunity to observe the latency,

duration and magnitude of time related changes between indicators. To date, no longitudinal studies have been found which describe or explain the effects these changing relationships may have on the interpretation of functional ability assessments during the rehabilitation period following knee injury.

While the limited information which is available regarding the criterion-related and construct validity of functional ability indicators following lower extremity injury has been derived from studies of injured knees, most of the data regarding evaluative (longitudinal construct) validity has been obtained from ankle injury studies. Airaksinen and associates (1990) measured ankle volume, perceived disability, pain, and ROM at intervals of 24 hours, 1 week, and 4 weeks post ankle sprain in a general orthopedic population (N = 44). Disability and pain were assessed with visual analog scales, swelling with volumetric displacement, and ROM with goniometry in a between groups study designed to demonstrate the effects of compression therapy on multiple dependent variables. Following baseline measurements of impairment and functional limitation, all indicators appeared to return toward normal values over time. However, this impression is based on graphic presentation of between groups data only. Analysis of within

subjects data for univariate or multivariate dependent measures was not accomplished, and duration of disability was not measured in this study.

Pennington and colleagues (1993) studied the effects of a single exposure to high-frequency electromagnetic fields on ankle swelling, range of motion, pain, and weight bearing ability among 50 military subjects with acute grade I and II ankle inversion sprains. The authors reported that swelling was reduced an average of 44 ml among treated subjects compared to 11 ml among sham-treated controls, resulting in a significant ($p=0.01$) one-tailed t-test. No inferential statistics were reported for the remaining dependent variables, and no attempts to correct for potentially inflated type I error rates among time-dependent covariates. Under the circumstances, the generalizability of the investigators' conclusion that treatment was successful in reducing ankle swelling and pain occurring less than 72 hours post-inversion sprain is questionable at best. However, they correctly acknowledged that decreased swelling in response to treatment has not been shown to speed recovery, and suggested that future studies should examine the relationships between physical measures and return to normal function following ankle sprain.

Wilkinson and Horn-Kingery (1993) did study the time-related changes in functional ability following acute grade I and II ankle inversion sprains. The elapsed time between ankle inversion injury, subjects' ability to perform nine different closed chain motor activities, including return to sports participation, was noted among 42 military academy cadets. A strong linear relationship was found between a hierarchical arrangement of the motor activities and days elapsed post injury (r=0.96-0.99 among three different treatment groups). However, the authors failed to assess the internal consistency of this hierarchical multi-item scale.

Data from the previous studies suggest that measures of motor activity, ROM, swelling, pain, and perceived functional ability are all sensitive to time-dependent change during rehabilitation following ankle sprain. However, more rigorous evaluation of the relative responsiveness and stability of these time-dependent variables would help to complete an assessment of their usefulness as evaluative measures in sports rehabilitation settings.

Kirshner and Guyatt (1985) state that the validity of clinical progress indicators is based on the parallel relationship between longitudinal changes in the indicator variables and the underlying trait of interest. They call

this evaluational validity or longitudinal construct validity. The point raised concerning parallelism, the relative rates of change between indicators, is an interesting notion which deserves further comment. We have already seen that indicators of patient progress should be sensitive to changes occurring throughout the entire rehabilitation period. For example, a poor measure may be sensitive to changes which occur early in rehabilitation, while displaying insensitivity to changes occurring later on. This so-called "ceiling effect" would render this measure useless as a source of information concerning clinical change after the initial responsive period. For this reason, Kirshner and Guyatt (1985) state that tests of time-related covariance should be included as indicators of longitudinal construct validity when evaluating the usefulness of clinical measures.

Summary

Disability is a complex concept encompassing physiological, behavioral, and environmental factors. This complexity is reflected in the sports medicine literature, which is rich with examples of candidate indicators and predictors of disability. However, neither the relationships between these measures within a conceptual model of

functional ability (construct validity) nor or their correlations with accepted indicators of treatment outcome (criterion-referenced validity) have been clearly established. Consequently, it has been impossible to adequately determine the extent to which the any of the vast array of dependent measures employed to detect treatment effects in sports medicine research relate to an athlete's demonstrated ability to perform following lower extremity injury.

Recognizing the need for theoretical and practical criteria by which to judge the effectiveness of ACL reconstruction procedures, several eminent practitioners from well-known sports medicine facilities published multi-item knee assessment instruments. However, the developers of these instruments lacked either the motivation or resources required to refine the content and scoring procedures of their "knee scores". As a result, researchers and clinicians have employed these measures as indicators of rehabilitation progress and treatment outcome without the benefit of critical assessment regarding either their scientific quality or suitability for the purpose at hand. In fairness, it should be pointed out that accepted standards for evaluating the validity of functional tests and measurements have been developed only recently (Rothstein, et al., 1991;

Johnston, et al., 1992). However, even these validity criteria have not been consistently or widely applied to dependent measures used in more recent studies appearing in the sports medicine literature. Consequently, it remains unclear which measures should be included in a comprehensive and accurate assessment of disability following lower extremity injury among athletes.

To address this problem, a conceptual framework based on Nagi's disablement model was introduced here to define the concept of disability, identify important components of that concept, and evaluate the relationships between those component domains in sports medicine settings. The content of the previously mentioned composite "knee scores" was examined in an attempt to find a concensus regarding assessment items which could serve as operational indicators of Nagi's disablement domains: impairment, functional limitation, and disability. Pain, swelling, and range of motion were found to be measures commonly used to assess physical impairment in these instruments. All of the composite instruments reviewed also included measures of observed or self-reported motor activities representing the domain of functional limitation. An operationalized functional ability model containing these measures was

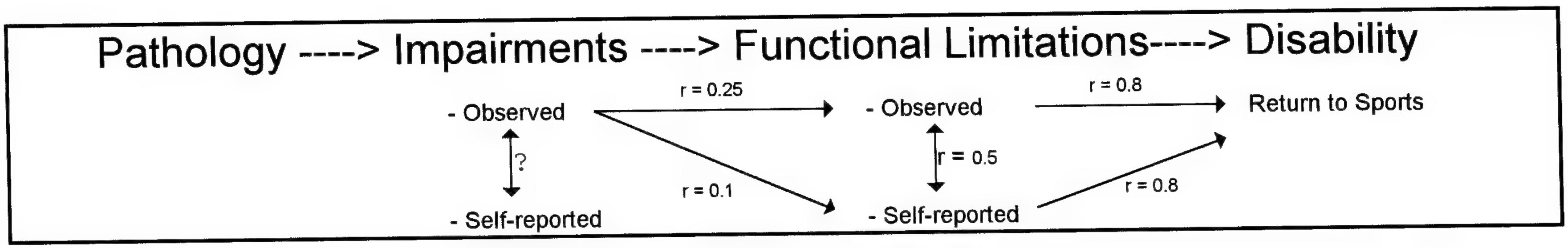

Figure 3. Summary of Relationships (median values of r) Between Disablement Indicators Reported in the Sports Medicine Literature

applied to data from the existing sports medicine literature (Figure 3). The relationships observed between measures of organic impairment, functional limitation, and disability levels suggest that functional limitation measures may be more useful disability measures than impairment measures. However, contextual problems inherent in the long-term follow-up of physical activity patterns have been implicated as potential sources of confounding measurement error in these studies.

All of the studies relating clinical indicators to functional outcomes have examined data gathered from knee injured subjects approximately two to six years following injury or surgical repair of the cruciate ligaments. These long-term follow-up studies pose particular measurement problems for the investigator, especially when the outcome criterion is athletic participation at a level which existed prior to injury. As Noyes (1989) observed, individuals may change sports activity levels due to life-style changes, graduation, changes in interest, or several other reasons not related to injury. In this context, intrasubject differences in potentially useful clinical indicators of impairment could be overwhelmed by environmental changes as significant sources of variance in functional ability. The greater the time elapsed since injury, the greater the

likelihood that these non-injury related factors may introduce themselves into a study as confounding variables. Studies aimed at clarifying the relationships between measures of functional ability in a rehabilitation setting should be conducted under conditions which are free of such unintended contextual contaminants.

Ankle sprains are the most common lower extremity injuries among athletes, and the recovery period following these injuries are relatively short. Consequently, virtually all of the longitudinal studies concerning the responsiveness and stability of measures (evaluative validity) observed during rehabilitation have been conducted on subjects with ankle sprains. As with the studies performed using knee-injured subjects, a review of the literature revealed that pain, swelling, and range of motion measures were the "common denominators" of organic dysfunction, and functional limitation has been measured using both self-reports and observations of motor activities. However, no studies addressing issues of construct or criterion-related validity of these measures following ankle sprains were found in the literature. Therefore, it is not known whether the weak relationships between organic impairment and disability measures documented among knee injured patients may be generalized to

athletes with other lower extremity injuries, including ankle sprains.

The weak correlations found between organic dysfunction and disability measures in sports medicine studies are especially problematic because most rehabilitation research employs dependent measures of anatomic abnormality or physiologic dysfunction to demonstrate the effects of treatment. Mechanically instrumented measures of impairment are generally regarded as trustworthy evidence of dysfunction, while symptomatic measures are rarely quantified or given adequate consideration in clinical decision making or research. Perhaps our reluctance to develop and use "subjective" measures to predict disability is an expression of the belief that observational and self-reported information is somehow unscientific. However, Johnston (1991) states that whether an instrument is a "subjective" observational rating or self-report rather than an "objective" physical measure is not necessarily a relevant scientific distinction. A more meaningful approach to instrument selection is based on an assessment of a measure's demonstrated validity and usefulness.

Therefore, to evaluate their usefulness as indicators and predictors of disability, the validity of impairment and functional limitation measures were compared within a study

of disability duration following acute ankle sprains among athletes. Measures representing both of these disability domains were selected based on their frequency of appearance in the sports medicine literature and were subsequently examined for evidence of construct validity, criterion-related validity, and evaluative (longitudinal construct) validity. Whether "subjective" behavioral measures of functional limitation are more or less accurate predictors of disability duration than "objective" physical measures of impairment is a related empirical question which is also a testable research hypothesis within the context of this validity study.

CHAPTER THREE

METHODOLOGY

Subjects

Twenty-one NCAA Division I athletes of both genders with recent Grade I and II ankle inversion sprains were used in this study. Informed consent to participate was obtained according to guidelines issued by the University of Virginia Committee for the Protection of Human Subjects (Appendix A). The subjects were enrolled in the study within 48 hours of injury and were asked to continue their participation until they were able to return to full athletic participation.

Procedures

Design

A multivariate within-subjects design was employed to assess the relationships between variables during the clinical rehabilitation period following acute ankle sprain. The clinical rehabilitation period was defined as the interval between the date of injury and return to athletic participation without limitations. Initial measurements were made within 72 hours from the time of injury. Impairment measures were repeated on each of the next two days in order

to establish the stability of baseline values. Subsequent measurement sessions were conducted at intervals of approximately one week until the subjects reported the ability to return to full athletic participation without limitation. On one occasion during the baseline period and again on each subsequent weekly measurement session, subjects were given a functional ability rating based on the number and type of closed chain activities they could perform on that day. Other variables included physical measures as well as measures of the subject's perceived level of disability and symptomatic responses to closed chain motor activity. Multiple measurements of each variable were taken to determine intraoccasion test-retest reliability.

Impairment Measures

Joint Swelling. Swelling of the affected ankle and foot was evaluated by the water displacement method. After removing socks and shoes, subjects were measured while sitting with the knee and hip flexed to approximately 90 degrees. A thin coat of water was applied to the limb prior to immersion to minimize the amount of air trapped around leg hair. Volumetric displacement of both ankles was measured using a commercially available foot volumeter

(Smith & Nephew Rolyan, Menomonee Falls, WI) using water at a temperature 80-85 degrees Celsius. Subjects were instructed to gently lower the limb into the water until the foot rested comfortably on the bottom of the volumeter with the heel and calf positioned to touch its rear wall. The displaced volume of water was discharged through an overflow spout, captured in a basin, and transferred to a 1000 ml graduated cylinder for measurement. The displaced volume of water was recorded to the nearest 5 ml for each of 2 consecutive trials.

Range of Motion. Subjects were placed in the prone position with the knee flexed to approximately 90 degrees. Passive dorsiflexion and plantar flexion were recorded bilaterally according to American Academy of Orthopedic Surgeons (1965) manual of goniometric procedures. Specific anatomical landmarks employed were the lateral malleolus (axis of motion), lateral aspect of the fifth metatarsal head (distal arm) and an imaginary line between the lateral malleolus and the fibular head (proximal arm). Subjects were instructed to inform the investigator of any pain experienced during the measurement. If pain was experienced during dorsiflexion or plantar flexion, the indicated point of pain onset was used as the position of end range

measurement. If no pain was encountered, passive range of motion end points were established using light manual overpressure at the fourth and fifth metatarsal heads. Total range of sagittal plane motion of the ankle and foot (maximum dorsiflexion to maximum plantar flexion) was recorded to the nearest degree during each of 3 consecutive measurements.

Pain. Subjects were asked to stand shoeless, with full weight on the injured ankle. While standing in this position, subjects were presented with a visual analog pain scale consisting of a 100 mm line situated between two sets of polar descriptors ('no pain' and 'worst pain imaginable'). Subjects were asked to indicate their pain using a single vertical mark on the line. Pain was recorded to the nearest millimeter by measuring the distance from the zero point ('no pain') to the mark made by the subject. If a subject was unable to bear full weight on the injured extremity, the maximum value of pain ('worst pain imaginable', 100 mm) was recorded on the scale.

Functional Limitation Measures

Observations of closed kinetic chain motor ability were recorded once during the baseline period (48-96 hours following injury) and again during each weekly measurement

session. Subjects wore their usual athletic shoes during all activities, which were performed on a smooth concrete surface.

At the beginning of the session, subjects were presented with the list of the closed chain activities and received a demonstration of each task from the investigator. Following each demonstration, subjects were asked whether they could perform that activity comfortably. If the subject answered yes, he/she was invited to perform the activity. All subjects were instructed to discontinue a task immediately if they experienced symptoms of ankle pain or instability. A dichotomous system was employed to score the activity scale, with one point awarded for successful completion of a task and no points awarded if a task was not attempted or aborted.

40 meter ambulation. Subjects walked through two trials over a straight 40 meter course and were classified according to gait pattern: full weight bearing without crutches, 3 points; partial weight bearing with crutches, 2 points; or non-weight bearing with crutches, 1 point. Subjects who used a combination of partial and non-weight bearing patterns with crutches were scored as non-weight bearing subjects (1 point).

40 meter run. Subjects ran a straight 40 meter course during each of two trials.

Figure-8 run. Cones at each end of a 6 meter course were designated as circling points. Subjects ran a figure-8 path around the cones during each of two trials. Each trial consisted of two laps of the course, approximately 24 meters.

Single hop. Subjects stood on the injured limb, then hopped as far as possible on 2 successive trials, landing on the same limb in accordance with the description provided by Daniel and colleagues (1982).

Cross-over hop. This activity was performed over a 6 meter course with a 15 cm. wide marker on the ground along its length. Subjects hopped three consecutive times on the injured limb, crossing over the center strip marker with each hop.

Stairs Hop. Subjects hopped up and down a flight of 14 steps on the injured extremity.

Self-reported athletic ability. A questionnaire (Appendices B-D) containing items related to perceived effort, joint stability, pain, and global assessment of

functional ability was administered to the participants following each motor ability testing session. The purpose of this questionnaire was to assess the subject's response to vigorous closed kinetic chain motor activity. A visual analog scale was used to score responses to each item. Subjects were asked to indicate their responses by making a single vertical mark on a 100 mm. line situated between two polar descriptors ('minimum' and 'maximum').

Disability duration (dependent variable). Disability duration was determined by subtracting the date of injury from the date of the athlete's return to full participation in practice or competition, whichever occurred first.

Data Analyses

Construct Validity

A major question addressed in this study concerned construct validity, the theoretical pattern of relationships between variables in a conceptual model. Evidence for construct validity was derived from the examination of patterns indicating divergence and convergence between variables (Campbell & Fiske, 1959). Two analytical techniques, hierarchical multiple regression analysis and path analysis, were employed to compare the patterns of divergence and convergence observed among the data gathered

in this study with patterns predicted by the conceptual model.

Divergence. Nagi's model depicts disablement as a sequential linear flow process progressing through 4 levels of theoretical reduction. Proof of divergence in this study depended upon evidence that each level of dysfunction identified in the model was distinct from the preceeding one. If disablement is indeed a sequential process, then measures at each level of dysfunction should explain more variance in disability duration than was explained by the preceeding level. Hierarchical multiple regression was used to test the null hypothesis that the amount of variance in disability duration explained when data from functional limitation measures were added to estimates obtained using measures of impairment was equal to zero.

Convergence. Path analysis is a causal modeling system which was used to demonstrate patterns of convergence among the variables in this study. This technique is based on the construction of a qualitative diagram in which every included variable is connected to others by directional paths (arrows) designating their causal relationships (Wright, 1960). It is important to notice that the sequential nature of the disablement process described by Nagi implies that impairment affects disability only

indirectly. That is, disability may be explained exclusively by the direct effects of functional limitation measures while the effects of organic dysfunction on disability are indirect, being completely mediated by, or subsumed under, functional limitation. This is easily seen in Nagi's conceptual model, which displays no direct paths from impairment to disability. If the effects of impairment on disability are truly subsumed under functional limitation, then the proportion of variance in disability duration explained by functional limitation measures alone should exceed the total variance explained by all the predictor variables preceeding it in the model. A goodness of fit statistic (Q) calculated using path coefficients (Specht, 1975) was used to determine whether the pattern of convergence specified in Nagi's conceptual model fit the data from this study.

Predictive Validity

Predictive validity of three competing disablement models was determined by comparing the standard errors of their mean predicted values using multiple regression analyses. These statistics were obtained by regressing disability duration, expressed as the number of days from injury to return to full athletic participation, onto the

array of predictor variables comprising each of the three models (impairment only, functional limitation only , and all 5 independent variables). This technique was employed to answer the research questions concerning whether, 1) disability duration could be accurately predicted from impairment measures, and 2) addition of observed motor ability scores and subjects' perceived functional ability ratings improved prediction of disability duration beyond that afforded by measures of impairment alone.

Longitudinal Construct Validity (Evaluative Validity)

In order to be considered valid, each indicator variable in a disablement model should make stable contributions to the total estimate of disability across occasions of measurement. That is, all variables in the model should demonstrate stable covariances indicating parallel relationships between the variables over time, or longitudinal construct validity. Evidence for longitudinal construct validity was obtained by comparing standardized regression coefficients (beta weights) for each variable obtained from regression equations generated by data gathered on both measurement occasions.

Measurement Stability

Test-retest Reliability. Intraoccasion stability of each of the impairment and functional limitation measures previously identified in this chapter was examined using the ICC calculations established by Shrout and Fleiss (1979). ICC formula 2,1 was selected to represent intraoccasion stability of each of the measures in this study because: (1) each measurement trial was considered a random sample from a larger population of trials, and (2) each measure on each trial was derived from a single value rather than composite score (Shrout & Fleiss, 1979). Specifically, calculations of intratester reliability were made by inserting the mean square values obtained from repeated measures analyses of variance into the following formula:

$$ICC_{(2,1)} = \frac{BMS - EMS}{BMS + (k - 1)EMS + [k(TMS - EMS)] / N}$$

where, BMS = between subjects mean square

EMS = error (residual) mean square

TMS = trial mean square

k = number of trials

N = number of subjects

The test-retest reliability of 2 derived measures, ROMLOSS and SWELLING was also examined. ROMLOSS was computed by subtracting each of three successive range of motion measurements taken on the injured ankle from each of 3 consecutive measurements taken on the uninvolved ankle. SWELLING was computed by subtracting each of 2 consecutive measures of volume displacement taken from the uninvolved ankle from 2 consecutive measures taken from the injured one.

The precision of all measures used in this study was determined by calculating the standard error of measurement (SEM) according to the following formula:

$$SEM = SD\sqrt{1 - ICC}$$

Internal consistency. The motor activity score used in this study was determined by summing the scores obtained from individual motor tasks. This practice assumes the property of additivity is present in the individual item scores. That is, all items are assumed to measure the same phenomenon and to be free of error introduced by correlation with extraneous factors. To test this assumption, item-total correlations were computed to determine the extent to which scores from each task contributed to the total score. Cronbach's alpha was also computed to determine the extent

to which the total score was free from error introduced by the inclusion of irrelevant tasks in the total score.

Responsiveness

Preplanned contrasts employing multiple paired t-tests (2-tails) were used to determine whether the indicator variables ROMLOSS, SWELLING, ACTIVITY, and DISABILITY were sensitive to changes occurring between 2 occasions of measurement separated by approximately one week.

To be considered useful, indicator variables should also produce stable estimates of disability duration on different occasions of measurement. Evidence of estimate stability was obtained by comparing the mean predicted values for disability duration produced by data gathered from 2 occasions of measurement separated by 1 week. Mean predicted values of disability duration were obtained from separate multiple regression analyses conducted on data gathered during each measurement occasion. A paired t-test was also used to test the equality of these means.

A Bonferonni correction (alpha/k, k = number of planned contrasts) was used to control for inflation of the familywise Type I error rate (alpha = .05) resulting from multiple pairwise comparisons. This procedure determined that a corrected critical p value of 0.01 (0.05/5) was

required for rejection of each null hypothesis positing equality of paired means.

CHAPTER FOUR

RESULTS

Twenty-one college age (20.3 $\pm$ 1.7 years) athletes of both genders (13 males, 8 females) participated in this study. Initial measurements were taken approximately 3 days (67.8 $\pm$ 15.2 hours) post-injury. Data from a second measurement session held approximately 1 week later (6.42 $\pm$ 1.25 days) were also recorded and analyzed.

Reliability Study

Test-retest Reliability

Intraclass correlation coefficients and standard errors of measurement for the predictor variables are displayed in Table 12. Intraoccasion intrarater stability was judged to be good for all of the measures ($ICC_{2,1}$ = 0.85-0.99). Mean square values used to calculate the ICCs appear in the repeated measures ANOVA table located at Appendix E.

Internal Consistency of the Activity Scale

SPSS^x RELIABILITY was used to analyze the internal consistency of the multi-item motor activity scale. Two items contained in the original scale, non-weight bearing ambulation and full athletic participation, were found to

Table 12

Intraclass Correlation Coefficients and

Standard Errors of Measurement

	$ICC_{2,1}$	SEM
PAIN	0.90	3.7 mm
ROM	0.95	2.8 deg
ROMLOSS	0.88	4.0 deg
VOLUME	0.99	8.9 ml
SWELLING	0.95	11.5 ml
ATHABILITY	0.86	16.8 mm
ACTIVITY	1.00	

have no variability ($s^2=0$) and were deleted from subsequent analyses and scoring. Internal consistency of the revised activity scale, as indicated by Cronbach's alpha, was 0.88. Individual item-total correlations and calculated values for alpha resulting if the remaining items were deleted from the total score are shown in Table 13.

Validity Study

Construct Validity

Hierarchical multiple regression and path analysis were employed to determine whether the qualitative pattern of relationships between variables described in Nagi's disablement model were consistent with quantitative data obtained from this study. Prior to analysis, each of the variables used in this study was assigned to a conceptual domain within Nagi's disablement model as depicted in Figure 4.

Analysis was performed using SPSS^x REGRESSION using mean substitutions for missing values augmented by SPSS^x FREQUENCIES in evaluation of assumptions. Results of the evaluation of assumptions led to no transformations of the variables to reduce skewness in their distributions, reduce the number of outliers, or improve the normality, linearity, and homoskedasticity of residuals. With the use of a $p < 0.001$ criterion for Mahalanobis distance, no multivariate

Table 13

Internal Consistency of the Motor Activity Scale

	ITEM-TOTAL CORRELATION	ALPHA IF ITEM DELETED
PARTWB	.377	.892
FULLWB	.629	.873
40M RUN	.792	.855
FIGURE 8	.835	.850
ONEHOP	.815	.852
TRIPHOP	.815	.852
STAIRHOP	.689	.868
LIMITED	.202	.901
		ALPHA = 0.88

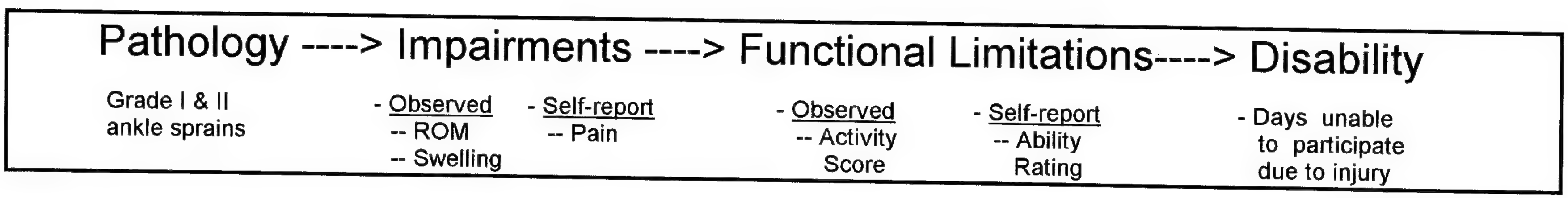

Figure 4. Operational Model of Disablement Following Acute Ankle Sprain

outliers were identified among the cases. Two cases had missing data for the predictor variable PAIN and no suppressor variables were found. N = 21. Table 14 displays the correlations between the variables, the unstandardized regression coefficients (B) and intercept, the standardized regression coefficients (Beta), the semipartial correlations (sr^2), and R, R^2 and adjusted R^2 after entry of all 5 independent variables. R was significantly different from zero at the end of each step. After step 2, with all independent variables in the equation, R = 0.828, F(5,15) = 0.82, p = 0.002. After step 1, with ROM loss, pain, and swelling in the equation, $R^2 = 0.490$, F_{inc} (3,17)= 5.45, p = 0.008. Addition of observed motor activity scores and self-reported athletic ability ratings to the equation resulted in a significant (p=0.027) increase in R^2. It was therefore concluded that functional limitation scores represented a phenomenon which was distinct from impairment, confirming the expected pattern of divergence between these theoretical levels of disablement as depicted in Nagi's conceptual model.

Convergence. A fully recursive path diagram (Figure 5) was used to represent all possible direct and indirect relationships between each of the variables in the

Table 14

Hierarchical Regression of Disability Duration on Impairment and Functional Limitation Variables

Variables	RETURN (DV)	PAIN	SWELLING	ROMLOSS	RATING	ACTIVITY	B	Beta	sr^2 (incremental)
PAIN	.316						.064	.106	
SWELLING	.383	-.148					.008	.068	
ROMLOSS	.457	-.026	.042				- .033	-.059	.490**
RATING	-.776	-.242	-.541	-.505			-.119	-.475	
ACTIVITY	-.718	-.261	-.194	-.699	.679		-1.056	-.395	.195*
						Intercept =	21.72		
Means	11.86	11.47	64.17	-9.03	47.81	4.90			
Std Dev	6.64	11.08	51.49	11.68	26.55	2.49			
								R^2	= 0.685***
								Adjusted R^2	= 0.580
								R	= 0.828

* p = 0.027
** p = 0.008
*** p = 0.002

conceptual model. The fully recursive path coefficients were developed from three regression analyses using SPSS^x REGRESSION with forced entry of the independent variables and mean substitution for missing values. First, activity level was regressed onto the package of three impairment variables and subjective athletic ability rating. Second, athletic ability rating scores were regressed onto a package consisting of the three impairment variables and activity level. Finally, disability duration was regressed onto all five independent variables.

Path coefficients from the residual terms (e_y) to their associated variables were calculated according to the formula developed by Wright (1934):

$$p = \sqrt{1 - R_y^2}$$

The fully recursive model accounted for the amount of total variance in disability duration ($R^2 = 0.685$) explained by all possible patterns of direct and indirect effects exerted by the 5 predictor variables included in the study.

In order to operationalize Nagi's theory accurately, the 3 paths representing direct effects of each of the impairment variables (PAIN, ROMLOSS, and SWELLING) on disability duration had to be deleted from the recursive model. This process was accomplished by setting these 3 path

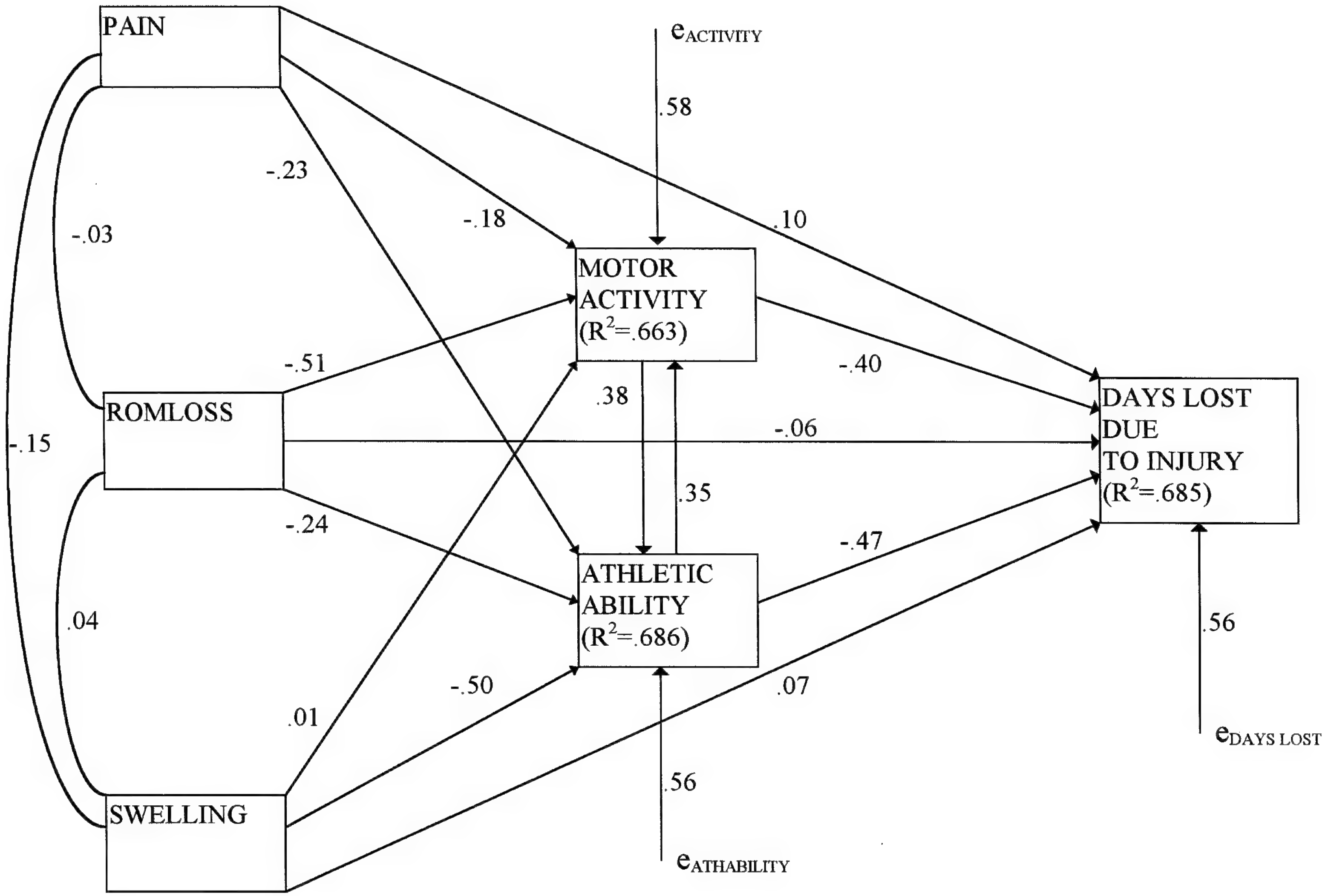

Figure 5. Fully Recursive Model with Path Coefficients

coefficients to zero, thereby creating an alternative or causal model (Figure 6) which is consistent with Nagi's disablement theory.

To determine whether the causal model representing Nagi's disablement theory fit the data from this study, the proportions of variance in days lost due to injury explained by both the recursive and causal models were compared. A goodness of fit statistic (Q) was used to determine the proportion of explained variance in disability duration which could be accounted for by the more restrictive causal model. The Q statistic was calculated using squared residual path coefficients from the fully recursive model in the numerator and those from the causal model in the denominator of the following expression (Specht, 1975):

$$Q = \frac{1 - R_m}{1 - M} = \frac{1 - \left[1 - (1 - R^2_{activity})(1 - R^2_{athability})(1 - R^2_{return})\right]}{1 - \left[1 - (1 - R^2_{activity})(1 - R^2_{athability})(1 - R^2_{return})\right]}$$

The calculated goodness of fit (Q = 0.977) indicated that a causal model based on Nagi's conceptual model of disability accounted for approximately 98% of the total explained variance in days lost due to injury experienced by the subjects. It was therefore concluded that the proposed qualitative disablement model fit the quantitative data obtained from the athletes in this study.

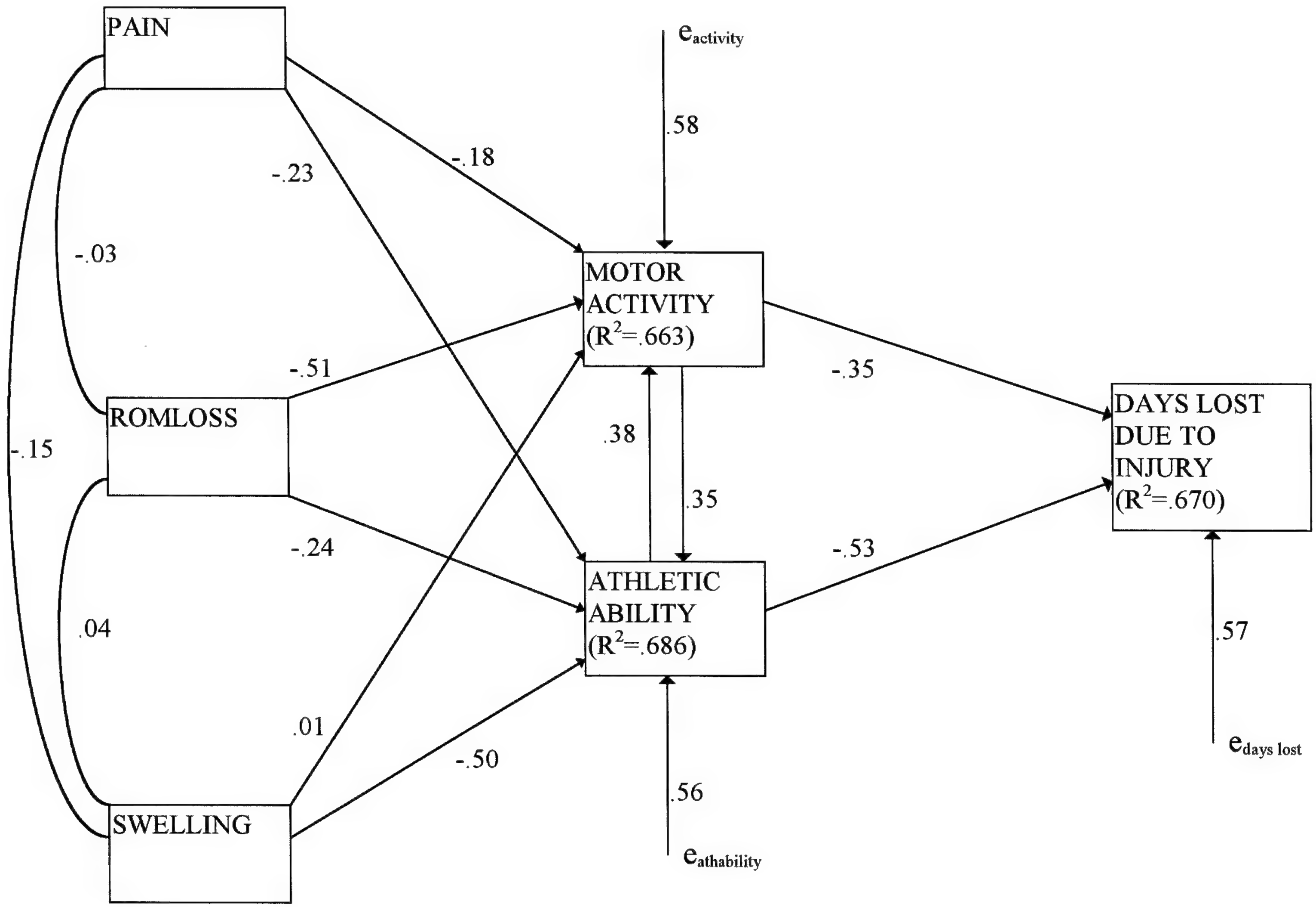

Figure 6. Causal Model with Path Coefficients

Predictive Validity

Predictive accuracy of the competing regression models was evaluated by comparing the standard errors of the mean predicted values for disability duration obtained using impairment, functional limitation, and all 5 independent variables (Table 15). The most accurate prediction (SE prediction =1.48) was derived from a regression equation containing only the functional limitation measures, motor activity scores and perceived athletic ability. Figure 7 illustrates the 95% confidence intervals around the mean predicted values obtained using the functional limitation measures as predictor variables.

Responsiveness

Preplanned contrasts employing multiple paired t-tests (2-tails) were used to determine whether the indicator variables ROMLOSS, SWELLING, ACTIVITY, and DISABILITY were sensitive to changes occurring during the rehabilitation period. A fifth t-test for paired samples was used to determine whether predictions of days lost due to injury were stable across 2 occasions of measurement separated by approximately 1 week. Results of the pairwise comparisons are summarized in Table 16. Using Bonferonni corrections to control Type I error rates among multiple paired t-tests,

Table 15

Accuracy of Mean Predicted Values of Disability Duration

Variables in Model (df)	Mean Predicted Value(days lost)	S E Prediction (days lost)	Lower Limit	Upper Limit
Impairment (17)	11.86	2.13	1.27	3.99
Functional (18)	11.86	1.48	1.03	2.22
All (15)	11.86	2.23	1.46	3.40

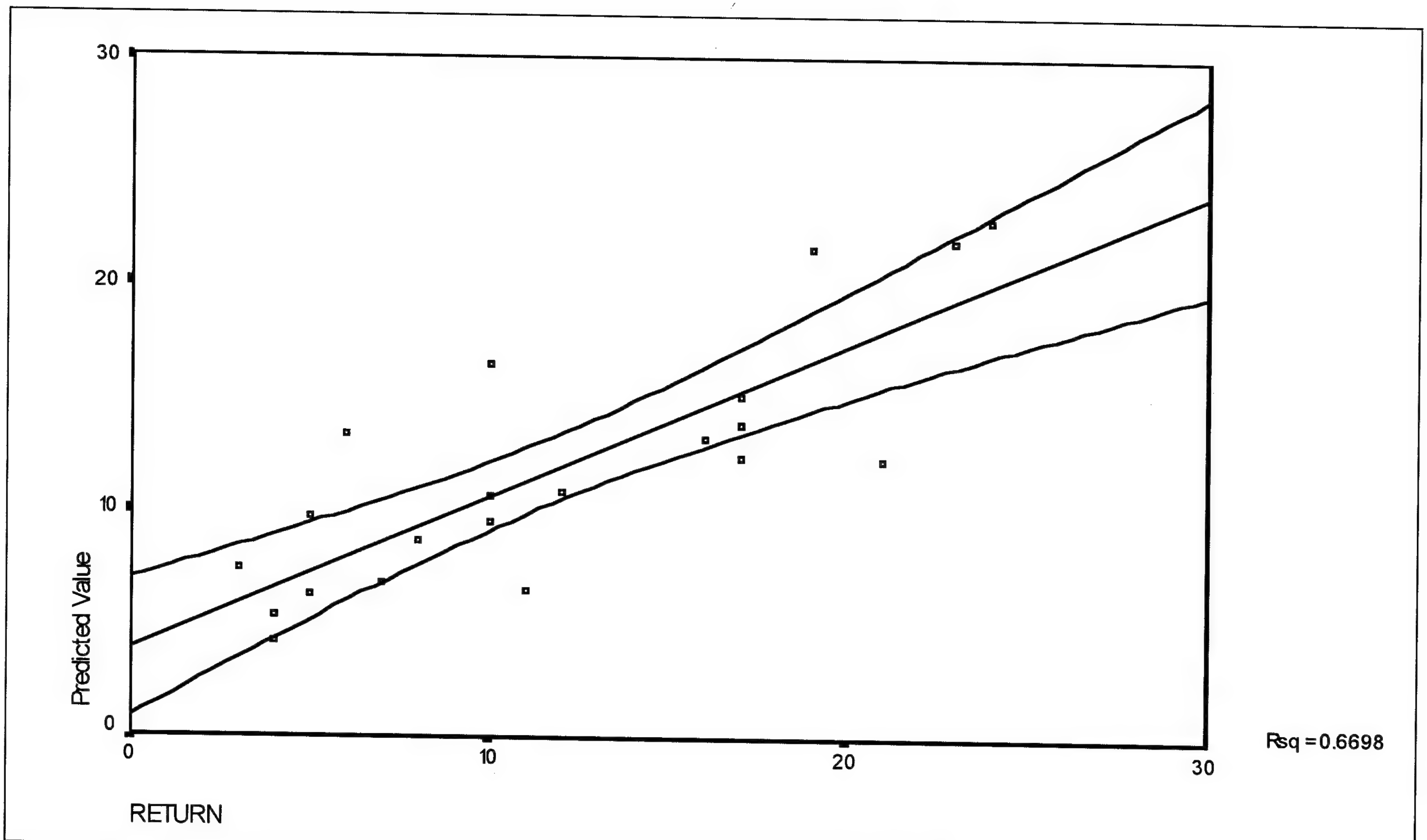

Figure 7. Plot of Observed and Predicted Disability Duration (days) Showing 95% Confidence Intervals Around Mean Predicted Values

Table 16

Responsiveness of Impairment and Functional Limitation Measures: Results of Paired t-tests (N=13)

Variable	Mean	SD	df	t-value	2-tail signif (alpha = .01)
Athletic Ability I	42.6	28.9			
Athletic Ability 2	72.2	31.6	12	-5.44	<.001*
ROMLOSS I	-12.3	12.7			
ROMLOSS 2	-7.6	9.2	12	-1.68	.119
Activity Score I	5.0	2.6			
Activity Score 2	6.9	1.9	12	-3.85	.002*
Swelling I	65.8	53.8			
Swelling 2	23.1	26.1	12	3.66	.003*

changes were detected within 3 of the 4 indicator variables: swelling, motor activity, and perceived athletic ability. The fourth measure, range of motion, increased by an average of 5° during the week between measurements, but this difference could not be distinguished from unexplained error (SD = 10°).

Stability of disability duration estimates. SPSS^x REGRESSION was used to determine whether predicted values for disability duration were stable across occasions of measurement. This procedure was employed twice, using the predictor variables ROMLOSS, SWELLING, ACTIVITY, and ABILITY from the baseline period to generate a disability prediction on occasion 1, and values obtained from the one week follow-up period to make a disability prediction on occasion 2. Mean predicted values of disability duration for both occasions were determined using SPSS^x REGRESSION with forced entry of all predictor variables. The mean predicted values from both occasions were then used to test the null hypothesis that a measurement interval of approximately 1 week would have no effect on predicted disability duration (Table 17). A calculated paired t-test value (2-tailed, df = 12) of 0.15 (p =.886) led to a conclusion that the null hypothesis could not be rejected. It was therefore concluded that disability duration predictions generated by the causal

model were unaffected by changes in the individual predictor variables occuring over a 1 week interval.

Longitudinal Construct Validity (Evaluative Validity)

Evidence for longitudinal construct validity was obtained by comparing standardized regression coefficients (beta weights) obtained for ROMLOSS, swelling, motor activity score, and perceived athletic ability on both occasions of measurement. These values are shown as percentages of the total explained variance in disability duration in Figure 8. These data led to the conclusion that the candidate indicators provided stable contributions to the total estimate of disability duration on both occasions, indicating parallel relationships between the variables over time, or longitudinal construct validity.

Table 17

Interoccasion Stability of Mean Predicted Values of Disability Duration: Results of Paired t-test (N=13)

Mean Predicted Value	Mean	SD	SEM	t-value	df	2-tail sig
Occasion 1	13.7	5.5	1.5			
Occasion 2	13.6	5.9	1.6	.15	12	.886

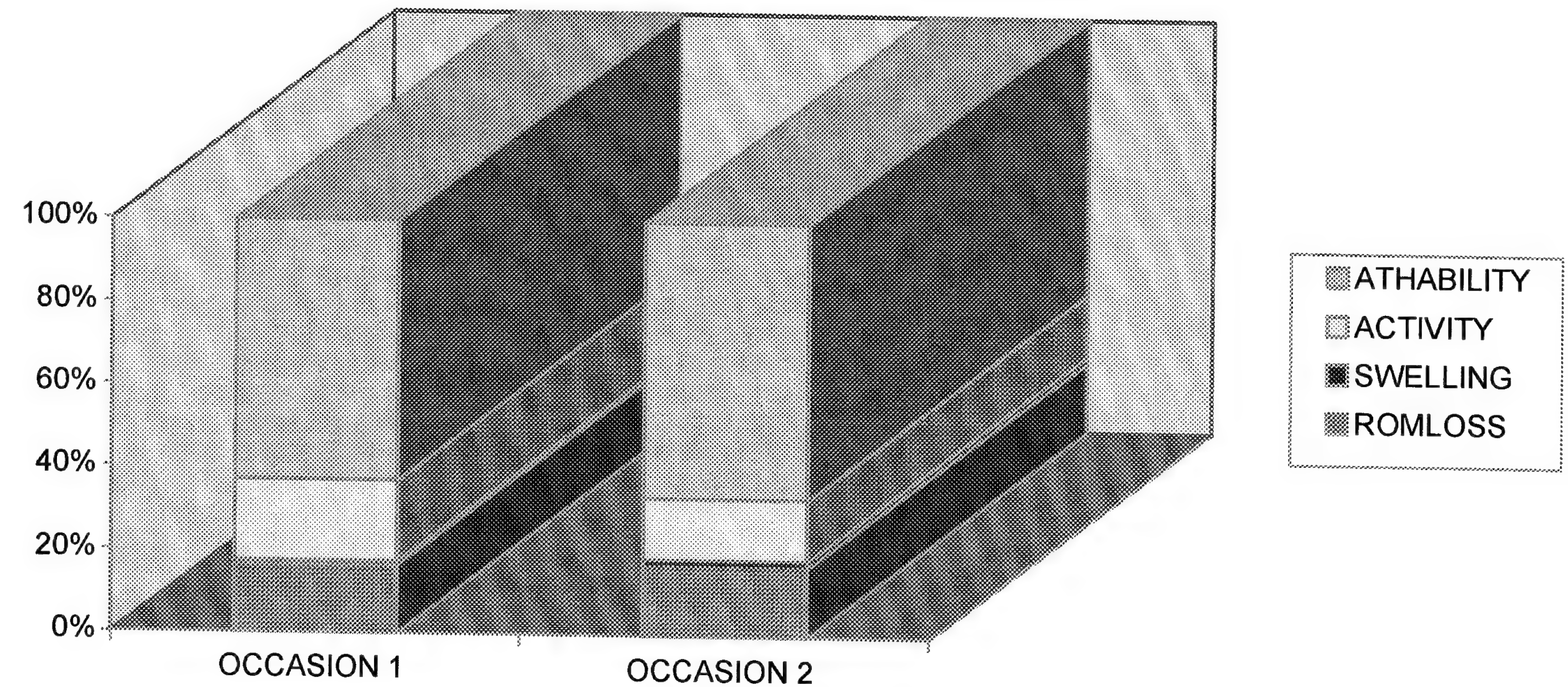

Figure 8. Longitudinal Construct Validity: Standardized Regression Coefficients Expressed as Percentages of Explained Variance in Disability Duration on 2 Occasions of Measurement

CHAPTER FIVE

DISCUSSION

This investigation was conducted to study the measurement validity of candidate variables representing organic and behavioral levels of dysfunction as indicators and predictors of disability observed among athletes following acute ankle sprain. The major findings of this study were, (1) a qualitative disability model developed for a general population fit the quantitative data obtained from a sample of NCAA Division I athletes following acute ankle sprain (constuct validity); (2) the duration of disability experienced by these athletes could be more accurately predicted from functional limitation measures than from impairment measures (criterion-referenced validity); (3) the functional limitation measures studied were responsive to time-related changes in patient condition while demonstrating patterns of covariance which were remarkably stable across an interoccasion measurement interval of about one week (longitudinal construct validity); and, (4) test-retest reliability coefficients (ICCs) and internal consistency measures (Cronbach's alpha) calculated for behavioral measures were comparable to reliability coefficients obtained from physical measures.

Validity

Construct Validity

The concept of construct validity is concerned with describing the relationships between variables and determining whether these relationships conform to expectations derived from theory, experience, or previous research. A major finding of this study was that a causal theory of disability based Nagi's conceptual model fit the data gathered from NCAA Division I athletes with recent ankle sprains. Specifically, this finding supports the assumption that the effects of organic dysfunction on disability duration are mediated by behavioral factors, such as motivation or stoicism, following acute ankle sprains among athletes. This conclusion was based on evidence from path analysis confirming patterns of convergence among the variables which were consistent with relationships specified in the conceptual model.

Path analysis is sometimes referred to as "multiple regression with pictures" because of its ability to graphically represent both the direct and indirect effects between variables through the use of path coefficients. The path coefficients in this study paint an interesting picture of the disablement process following acute ankle sprain

among athletes. The weak relationships observed between ROM and swelling were unexpected and may be used to illustrate a point concerning precision, reliability, and validity of measures.

Volumetric measures have been repeatedly shown to be precise and consistent measures of swelling. These properties concern measurement reliability. However, swelling may occur as result of a number of different causes, including jount effusion and soft tissue edema accompanying gravity-dependent positioning. Volumetry does not distinguish betwen joint effusion and edema in the surrounding tissues. The question concerning which phenomena are included in volumetric measurement is a validity issue. If, for example, loss of ROM is related only to joint effusion, then that portion of volumetry which measures tissue edema actually represents measurement error when the purpose of measurement is to predict ROM. This example illustrates the fact that precise and reliable instruments do not necessarily produce measures which are valid for a given purpose. Furthermore, measures with demonstrated validity for one particular purpose, such as indicating improvement in rehabilitation status (evaluative validity) may not prove valid for another purpose, such as prediction of treatment outcome.

Predictive Validity

Predictive validity is concerned with the accuracy of forecasts and requires comparing predicted outcomes with observed outcomes of treatment. In practice, predictive validity is actually the easiest type of validity to prove. All that is needed are a set of predictor variables, an outcome variable, patients (with similar conditions), and patience (Kirshner & Guyatt, 1985). Once outcome data have been gathered, multiple regression is used to yield a formula which weights the predictor variables in such a way that the amount of error in the predicted value is minimized. Observed values are then compared with predicted values to determine the accuracy, or predictive validity, of the variables used in the regression equation.

In this study, the outcome (criterion) measure was the number of days actually lost due to injury following Grade I or II ankle sprain. The mean value for observed disability duration (12 $\pm$ 6.6 days) agrees with data previously gathered by Wilkerson and Horn-Kingery (1993), who reported an average of 13 ($\pm$ 5.5) days lost due to injury among US Air Force Academy cadets with similar ankle injuries.

Three separate predictions based on, (1) impairment measures only, (2) functional limitation measures only, and

(3) all measures were compared for accuracy using the standard error of each mean predicted value. This statistic provides 95% confidence intervals around outcome predictions generated from multiple regression equations. Analysis revealed that the most accurate prediction of disability duration was based on functional limitation measures alone. This finding indicates that the impairment measures used in this study introduced more error into outcome predictions than the functional limitation measures.

In order to determine the source(s) of this predictive error, it would be helpful to have some knowledge concerning any underlying factors ("latent traits") contributing to estimates of disability duration. In this study, for example, motivation directed toward return to participation may influence time lost due to disability. Although none of the measures used to estimate disability duration was intended to assess motivation, scores produced by measures requiring effort or cooperation on the part of the subjects may reflect this underlying trait. If motivation were an underlying factor contributing substantially to disability duration, then voluntary measures reflecting the effort or motivation levels of the subjects (motor activity, perceived athletic ability, and ROMLOSS) could be expected to display some covariance with the outcome measure. Inclusion of

impairment measures requiring no effort, such as swelling, would introduce variance unrelated to motivation, resulting in increased error of prediction.

The existence of latent factors within measurement schemes is often confirmed using a multivariate statistical technique, factor analysis. However, the low subject to variable ratio in this study precludes the use of this analytical method in the present study.

Reliability

Reliability is an essential issue when evaluating the validity of functional ability indicators. All measurements contain some degree of error. The concept of reliability is concerned with the degree to which a measure is free of error, and the relevant form of reliability used to evaluate an instrument depends on the most likely type of error and the nature of the measure under consideration (Johnston et al., 1992).

Test-retest Reliability

Test-retest reliability, or intraoccasion stability, of scores is a basic characteristic needed to assess the quality of any clinical measure. Knowledge of stability over time is absolutely essential to establish the reliability of

self-report measures such as pain or perceived athletic ability because symptoms and other intrinsic phenomena are difficult to confirm or standardize between subjects and impossible to compare across observers. Test-retest reliability coefficients in the form of ICCs computed from "subjective" measures in this study demonstrate that pain and perceived functional ability may be quantified with relatively small degrees of error using visual analog scales. In fact, the test-retest reliability scores computed for self-reported measures of pain ($ICC_{2,1}$ = 0.91; SEM = 3.3 mm) and perceived athletic ability ($ICC_{2,1}$ = 0.86; SEM = 9.7 mm) used in this study were comparable to the reliability coefficients derived from "harder" physical measures.

The intraoccasion test-retest reliability calculated for ankle dorsiflexion-plantar flexion in the present study ($ICC_{2,1}$ = 0.95, SEM = 2.81°) is somewhat greater than those previously reported by Elveru and colleagues (1988)($ICC_{1,1}$ = 0.86-0.90). Three methodological differences may account for this. First, Elveru and associates computed ICCs using Shrout and Fleiss' formula 1,1 rather than formula 2,1 which was used in this study. Formula 1,1 is appropriate whenever successive trials do not contain scores for each subject or when each rater does not score each subject (Shrout & Fleiss, 1979). Second, each rater in Elveru's study was

allowed to choose the goniometric technique and patient position employed in the measurement. This absence of procedural control could be expected to introduce measurement techniques containing considerable measurement error into the data, resulting in lower overall intrasession stability. Finally, Elveru measured plantar flexion and dorsiflexion separately, while measurements from the present study reflect total sagittal plane motion at the ankle. The former method requires the observer to estimate a neutral reference point (0 degrees of dorsi/plantar flexion) as well as the endpoints of motion, introducing an additional opportunity for error to enter each measurement.

Volumetric measurements of ankle swelling have previously been shown to exhibit high degrees of precision, producing coefficients of variation (CV) in the 0.3-0.6% range (Smyth et al., 1963; Goldie et al., 1974). However, precision of measurements and stability of measurements are not identical concepts. Furthermore, coefficients of variation are of dubious value as indicators of reliability when between subjects variance exists within the characteristic being measured. This is because CVs do not differentiate between variance within the sample and measurement error (Rothstein & Echternach, 1993). Reliability coefficients computed from volumetric

displacement data in this study revealed a low proportion of measurement error to total error ($ICC_{2,1}$ = 0.99) while confirming the high degree of precision (SEM = 8.88 ml) reported for this measure in earlier studies.

Internal Consistency of Motor Activity Scores

Observational measures, such as the motor ability score used in this study, may be tainted by rater bias. Rater bias refers to a systematic variation leading to consistently higher or lower scores reported by different observers. The absence of bias, or objectivity, in a measure is commonly assessed using interrater reliability coefficients. The fact that interrater reliability data were not gathered here represents a weakness of this study.

However, Kirshner and Guyatt (1985) note that the interrater reliability of an observational measure is improved if scoring does not require interpretive judgment or conjecture on the part of the observers. In this regard, advantage accrues to the use of dichotomous scoring of well-defined observable phenomena such as motor task completion. This desirable characteristic minimizes error or "background noise", making it easier to observe any differences in functional status (Kirshner & Guyatt, 1985). Motor task completion was the scoring criterion used as a measure of

functional limitation in this study. When item responses consist of a rater's observations of performance, inconsistencies in scoring will adversely affect internal consistency of the total score. Although it is not an adequate substitute for interrater reliability coefficients, the internal consistency coefficient of motor activity scores in this study (Cronbach's alpha = 0.88) suggests that relatively little observational error is contained in this measure.

Instead of dichotomous measures of task completion, performance-related measures have been used as indicators of functional limitation in previous studies (Barber et al., 1990; Gauffin et al., 1990; Lephart et al., 1992). These performance measures may introduce between-subjects differences in agility, speed, leaping ability and other attributes of athletic ability as unexplained sources of variance when these scoring techniques are used to assess functional limitation. As a result, faster or stronger athletes will tend to appear to be less functionally limited than slower or weaker ones. The inclusion of irrelevant traits in performance-related measures may partially explain the relatively poor correlations found between impairment and functional limitation observed in the earlier studies.

Responsiveness of Measures

We have already mentioned that knowledge of a measure's intraoccasion stability is essential to determine reliability. The concept of stability is also a critical consideration when a measure is used to indicate changes in an athlete's functional status occurring during rehabilitation. To accurately measure these changes, it would be desirable to select measures which are sensitive to actual improvement in patient function over the period of interest while demonstrating little unexplained or random variation between measurements (Kazdin, 1982; Kirshner & Guyatt, 1985). That is, valid change measures must be both stable within occasions of measurement and responsive to actual changes in functional status occurring between occasions of measurement.

In this study, data obtained during 2 different observation occasions separated by a period of approximately one week were examined to determine the responsiveness of the indicator variables. Based on their demonstrated responsiveness to time-related change, swelling, motor activity, and perceived athletic ability appear to be potentially valid indicators of athletes' rehabilitation progress following ankle sprain. These findings support the

earlier work of Airaksinen and colleagues (1990), and Pasila and associates (1978). However, readers are reminded that the results of the current study are based on observations made at an interval of only one week between measurements. These candidate indicators of disability status may not all change at the same rates during other periods during rehabilitation, and their relative importance as indicators of functional ability may vary over longer courses of treatment.

Conclusions and Recommendations

In this study, a conceptual model of functional ability based on Nagi's disablement theory was applied to data obtained in a sports medicine setting. This conceptual model views disablement as a linear flow process which proceeds sequentially from pathology through organic impairment and functional limitation to disability. These domains represent different levels of theoretical reduction with regard to physical disability: organic, behavioral, and contextual. Based on the results of this study, reduction of Nagi's theory to address function primarily at the behavioral level rather than the organic level appears to best fit the practice of sports rehabilitation.

The level of function at which we, as rehabilitation professionals, choose to observe the effects of intervention is an important theoretical and practical question with implications for both clinical practice and research. As Michels (1994a) has noted, most observations in rehabilitation settings are directed at movement dysfunction and motor activities, not organs. From this behavioral perspective, movement and motor function become the directly observable phenomena from which organic function is inferred. This behavioral view of functional ability assessment is also central to the primary purpose of physical therapy:

> "...the promotion of optimal human health and function through the application of scientific principles to prevent, identify, assess, correct, or alleviate acute or prolonged movement dysfunction." (APTA Philosophical Statement on Physical Therapy, RC 5-83).

Focusing on dysfunction at the behavioral level for observation and measurement of disability holds the potential for development of science and theory unique to physical rehabilitation. Viewed from this perspective, the goal of physical rehabilitation research is to produce a body of knowledge about motor behavior that can then be used to plan and implement clinical procedures to modify dysfunctional motor behavior (Payton, 1993). The advantages

afforded by this level of theoretical reduction pose opportunities for postulating and testing relationships between and among motor activities, and between motor activities and criterion measures of athletic performance, for identifying organic variables which affect motor ability and manipulating those variables (if they are modifiable), and for drawing conclusions about motor function and putting those conclusions into practice.

Most rehabilitation research, including sports rehabilitation, has historically been oriented toward demonstrating treatment effects occurring at the tissue and organ level. In the research process, the concept of treatment effect becomes rather narrowly defined by observed changes in the investigator's dependent measure, and the investigator is statistically rewarded for selecting dependent measures which are sensitive to treatment. However, clinicians and investigators who mistakenly rely on statistical significance as the sole validity criterion in clinical research should recognize that changes in organic impairment accompanying treatment achieve clinical significance if, and only if, they are associated with observable changes in functional ability. Unfortunately, observed effects of treatment on organic function are seldom related to concomitant changes in observed motor function or

disability outcome. Instead, observed effects are implicitly linked to changes in function based on leaps of biomedical logic (e.g.- "Function requires strength, treatment X improves strength, therefore, treatment X improves function."). This practice has manifested itself in the regrettable notion that particular measures, whose relationships to disability have not been established, are adequate "tests" of an athlete's functional ability (Rothstein, 1994).

> "For example, there was a seductive quality to the isokinetic torque curve that could seemingly explain why a person could not walk or whether a person could return to athletic competition. We forgot that the torque curve should have been only a part of a complex story in which many variables are considered." (Rothstein, 1994)

My purpose in citing this example is not to discredit a particular form of measurement. Rather, it is to demonstrate the consequences of becoming too reliant on any single measure to represent all aspects of dysfunction in both research and clinical decision making. I do not advocate abandoning organic measures of treatment effects in sports rehabilitation research and practice, but I do advocate that we abandon our ignorance of so-called "subjective" measures which appropriately focus on function at the motor behavioral level.

Perhaps in the belief that behavioral information required to comprehensively assess function is too unreliable or "soft" for serious scientific consideration, clinicians and researchers have tended to minimize the weight given to a patient's symptoms when evaluating functional ability (Delitto, 1989). The result is that these responses to injury are rarely given adequate consideration during the research or clinical decision making processes. However, Rothstein (1989) has criticized the practice of minimizing information which is based on self-report or observational measures, arguing that any measure which is reliable, regardless of how it is obtained, is potentially useful in clinical decision making.

The data obtained from this study indicate that self-reports of pain and perceived athletic ability derived from visual analog scales contain small degrees of measurement error, varying little from one observation to another during a single occasion of measurement. Furthermore, observational measures of motor behavior and athletes' beliefs in their ability to perform were shown to provide more accurate predictions of the number of days lost due to injury following acute ankle sprain than precise physical measures such as ankle volumetry. Behavioral measures also demonstrated responsiveness to improvements in functional

status while displaying stable covariate relationships with physical measures during rehabilitation. Based on these findings, the notion that these demonstrably useful behavioral measures are somehow less than scientific should be discarded.

Payton (1993) points out that one of the primary tasks of researchers is to provide clinicians with valid tools which can then be used to measure therapeutic effects or future functional status. I strongly encourage future projects aimed at developing more sophisticated behavioral measures of motor dysfunction than the crude implements used here. Future studies should simultaneously test the relationships between behavioral measures, organic impairment, and functional outcome criteria in order to: (1) examine the validity of other organic and behavioral measures, and (2) determine whether effect sizes observed at the organic and behavioral levels in response to treatment result in meaningful reductions in days lost due to injury.

The simple conceptual model presented here is not the only theoretical scheme which may be applied to these data. We must continue to develop rehabilitation theory to ensure that valid clinical measures are used to relate therapeutic practices to clinical outcomes. As Michels (1994b) has stated, the consequences of developing trustworthy measures

of motor dysfunction, then simultaneously examining the relationships between measurable changes in functional outcome at the organic, behavioral, and contextual levels of theory reduction can potentially benefit both the science and practice of sports rehabilitation.

> "If we are to develop our own science for understanding and doing socially useful things about human movement dysfunction, that science must first be well grounded in the definition, identification, and study of relevant, observable phenomena and variables and later explained and elaborated by appeal to more abstract and organic levels of reduction. The robust sciences work forth and back between the nonabstract and the abstract levels of reduction. Our field persists in trying to work back from the abstract levels of reduction without working forth to the nonabstract levels of reduction, an intellectual practice that, by itself, constitutes searching for rationale, not creation of science." (Michels, 1994b)

REFERENCES

American Academy of Orthopaedic Surgeons (1965). Joint motion: Methods of measuring and recording. AAOS, Chicago.

Andersson, C., Odensten, M., Gillquist, J. (1991). Knee function after surgical or nonsurgical treatment of the anterior cruciate ligament: A randomized study with a long-term follow-up period. Clinical Orthopaedics and Related Research, 264, 255-262.

Airaksinen, O., Kolari, P.J., and Miettinen, H. (1990). Elastic bandages and intermittent pneumatic compression for treatment of acute ankle sprains. Archives of Physical Medicine and Rehabilitation, 71, 380-383.

Arnold, J.A., Coker, T.P., Heaton, L.M., Park, J.P., Harris, W.D. (1979). Natural history of anterior cruciate tears. American Journal of Sports Medicine, 7, 305-313.

Asher, H.B. (1976). Causal Modeling. Sage Publications, Beverly Hills, CA.

Babbie, E. (1990). Survey Research Methods, 2nd edition. Wentworth Publishing Co., Wentworth, CA.

Barber, S.D., Noyes, F.R., Mangine, R.E., McCloskey, J.W., Hartman, W. (1990). Quantitative assessment of functional limitations in normal and anterior cruciate ligament-deficient knees. Clinical Orthopaedics and Related Research, 255, 204-214.

Boneau, C.A. (1960). The effects of violations of assumptions underlying the t test. Psychological Bulletin, 37, 49-64.

Campbell, D.T. and Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Cook, T.D. and Campbell, D.T. (1979). Quasi-experimentation: Design & Analysis Issues for Field Settings, Houghton Mifflin Co., Boston.

Curl, W.W., Markey, K.L., and Mitchell, W.A. (1983). Agility training following anterior cruciate reconstruction. Clinical Orthopaedics and Related Research, 172, 133-136.

Daniel, D.M., Malcolm, L., Stone, M.L., Perth, H., Morgan, J., Riehl, B. (1982). Quantification of knee stability and function. Contemporary Orthopaedics, 5, 83-91.

Daniel, D.M., Stone, M.L., Riehl, B., Moore, M.R. (1988). A measure of lower limb function. American Journal of Knee Surgery, 212-214.

Delitto, A. (1989). Subjective measures in clinical decision making. Physical Therapy, 69, 585-589.

Elveru, R.A., Rothstein, J.M., and Lamb, R.L. (1988). Goniometric reliability in a clinical setting. Subtalar and ankle joint measurements. Physical Therapy, 68, 672-677.

Feinstein, A.R. (1987). Clinimetrics, Yale University Press, New Haven.

Flandry, F., Hunt, J.P., Terry, G.C., and Hughston, J.C. (1991). Analysis of subjective knee complaints using visual analog scales. American Journal of Sports Medicine, 19, 112-118.

Gauffin, H., Petterson, G., Tegner, Y., Tropp, H. (1990). Function testing in patients with old rupture of the anterior cruciate ligament. International Journal of Sports Medicine, 11, 73-77.

Goldie, I., Gunterberg, B., and Jacobson, C. (1974). Foot volumetry as an objective test of the effect of antiphlogistic drugs in ankle sprains. Rheumatology and Rehabilitation, 13, 204-207.

Harter, R.A., Osternig, L.R., Singer, K.M., James, S.L., Larson, R.L., and Jones, D.C. (1988). Long-term evaluation of knee stability and function following surgical reconstruction for anterior cruciate ligament insufficiency. American Journal of Sports Medicine, 16, 434-443.

Heerkens, Y.F., Brandsma, J.W., Lakerveld-Heyl, K. and van Ravensberg, S.D.(1994). Impairments and disabilities--the difference: proposal for adjustment of the International Classification of Impairments, Disabilities, and Handicaps. Physical Therapy, 74, 430-442.

Hocutt, J.E. Jr., Jaffe, R., Rylander, C.R., and Beebe, J.K. (1982). Cryotherapy in ankle sprains. American Journal of Sports Medicine, 10, 316-319.

Hulin, C.L., Drasgow, F., and Parsons, C.K. (1983). Item Response Theory: Application to Psychological Measurement, Dow Jones-Irwin, Homewood, IL.

Hutchinson, T.A., Boyd, N.F., and Feinstein, A.R. (1979). Scientific problems in clinical scales, as demonstrated by the Karnofsky index of performance status. Journal of Chronic Disease, 32, 661-666.

Jensen, J.E., Slocum, D.B., Larson, R.L., James, S.L., and Singer, K.M. (1983). Reconstruction procedures for anterior cruciate ligament insufficiency: a computer analysis of clinical results. American Journal of Sports Medicine, 11, 240-8.

Jette, A.M. (1994). Physical disablement concepts for physical therapy research and practice. Physical Therapy, 74, 380-386.

Johnston, M.V., Findley, T.W., DeLuca, J., and Katz, R.T. (1991). Research in physical medicine and rehabilitation XII: measurement tools with application to brain injury. American Journal of Rehabilitation Medicine, 70, 40-56.

Johnston, M.V., Kieth, R.A., and Hinderer, S.R. (1992). Measurement standards for interdisciplinary medical rehabilitation. Archives of Physical Medicine and Rehabilitation, 73, S3-S23.

Kazdin, A.E. (1982). Single-case research designs: Methods for clinical and applied settings, New York, NY, Oxford University Press, Inc.

Keller, P.M., Shelbourne, K.D., McCarroll, J.R., and Rettig, A.C. (1993). Nonoperatively treated isolated posterior cruciate ligament injuries. American Journal of Sports Medicine, 21, 132-136.

Kettelkamp, D.B. and Thompson, C. (1975). Development of a knee scoring scale. Clinical Orthopaedics, 107, 93-99.

Kirshner, B. and Guyatt, G. (1985). A methodological framework for assessing health indices. Journal of Chronic Disease, 38, 27-36.

Lankhorst, G.J., Van de Stadt, R.J., Van der Korst, J.K. (1985). The relationships of functional capacity, pain, and isometric and isokinetic torque in osteoarthrosis of the knee. Scandinavian Journal of Rehabilitation Medicine, 17, 167-72.

Law, M. (1993). Evaluating activities of daily living: Directions for the future. American Journal of Occupational Therapy, 47, 233-237.

Lephart, S.M. (1991). Doctoral dissertation, University of Virginia, Charlottesville, VA.

Lephart, S.M., Perrin, D.H., Fu, F.H., Gieck, J.H., McCue III, F.C., and Irrgang, J.J. (1992). Relationship between selected physical characteristics and functional capacity in the anterior cruciate ligament-insufficient athlete. Journal of Orthopaedic and Sports Physical Therapy, 16, 174-181.

Linde, F., Hvass, I., Jurgensen, U., and Madsen, F. (1984): Compression bandage in the treatment of ankle sprains. A comparative prospective study. Scandinavian Journal of Rehabilitation Medicine, 16, 177-9.

Lysholm, J. and Gillquist, J. (1982). Evaluation of knee ligament surgery results with special emphasis on use of a scoring scale. American Journal of Sports Medicine, 10, 150-154.

Marshall, J.L., Fetto, J.F., and Botero, P.M. (1977). Knee ligament injuries: A standardized evaluation method. Clinical Orthopaedics & Related Research, 123, 115-29.

Maxwell, C. (1978). Sensitivity and accuracy of the visual analog scale. British Journal of Clinical Pharmacology, 6, 15-24.

Michels, E. (1994a). Movement dysfunction: Meaning and use of the term. American Physical Therapy Section on Research Newsletter, 27(2), 1-4.

Michels, E. (1994b). Levels of reduction (letter), Physical Therapy, 74, 1138-1139.

Nagi, S. (1991). Disability concepts revisited: Implications for prevention. In Pope, A.M. and Tarlov, A.R. (editors): Disability in America: Toward a National Agenda for Prevention. National Academy Press, Washington, D.C., 309-327.

Noyes, F.R., Barber, S.D., and Mangine, R.E. (1990). Bone-patellar ligament-bone and fascia lata allografts for reconstruction of the anterior cruciate ligament. Journal of Bone & Joint Surgery, 72-A, 1125-36.

Noyes, F.R., Barber, S.D., and Mangine, R.E. (1991). Abnormal lower limb symmetry determined by function hop tests after anterior cruciate ligament rupture. American Journal of Sports Medicine, 19, 513-518.

Noyes, F.R., Barber, S.D., and Mooar, L.A. (1989). A rationale for assessing sports activity levels and limitations in knee disorders. Clinical Orthopaedics & Related Research, 246, 238-249.

Noyes, F.R., Grood, E.S., Butler, D.L., Malek, M. (1980). Clinical laxity tests and functional stability of the

knee: Biomechanical concepts. Clinical Orthopaedics & Related Research, 146, 84-49.

Noyes, F.R., Matthews, D.S., Mooar, L.A., and Grood, E.S. (1983). The symptomatic anterior cruciate deficient knee. Part II: The result of rehabilitation, activity modification, and counseling on functional disability. Journal of Bone and Joint Surgery, 65-A, 163-174.

Noyes, F.R., McGinniss, G.H., and Mooar, L.A. (1984). Functional disability in the anterior cruciate deficient knee syndrome. Review of knee rating systems and projected risk factors in determining treatment. Sports Medicine, 1, 278-302.

Noyes, F.R. and McGinniss, G.H. (1985). Controversy about treatment of the knee with anterior cruciate laxity. Clinical Orthopaedics & Related Research, 198, 61-76.

Odensten, M., Lysholm, J., Gillquist, J. (1983a). Long-term follow up study of a distal iliotibial band transfer (DIT) for anterior cruciate instability. Clinical Orthopaedics & Related Research, 176, 129-135.

Odensten, M., Tegner, Y., Lysholm, J., Gillquist, J. (1983b). Knee function and muscle strength following distal ileotibial band transfer for anterolateral rotatory instability. Acta Orthopaedica Scandinavica, 54, 924-928.

Parolie, J.M. and Bergfeld, J.A. (1986). Long-term results of nonoperative treatment of isolated posterior cruciate injuries in the athlete. American Journal of Sports Medicine, 14, 35-38.

Payton, O.D.(1993). Measurement validity in physical therapy research (commentary). Physical Therapy, 73, 114-115.

Pennington, G.M., Danley, D.L., Sumko, M.H., Bucknell, A., and Nelson, J.H. (1993). Pulsed, non-thermal, high-frequency electromagnetic energy (DIAPULSE) in the treatment of grade I and grade II ankle sprains. Military Medicine, 158, 101-104.

Pasila, M., Visuri, T., and Sundholm, A. (1978). Pulsating shortwave diathermy: Value in treatment of recent foot and ankle sprains. Archives of Physical Medicine and Rehabilitation, 59, 383-386.

Pope, A.M. and Tarlov, A.R. (editors), (1991): Disability in America: Toward a National Agenda for Prevention. National Academy Press, Washington, D.C., 76-108.

Price, D.D., McGrath, P.A., Rafii, A., and Buckingham, B. (1983). The validation of visual analog scales as ratio scale measures for chronic and experimental pain. Pain, 17, 45-54.

Risberg, M.A. and Ekeland, A. (1994). Assessment of functional tests after anterior cruciate ligament surgery. Journal of Orthopaedic and Sports Physical Therapy, 19, 212-217.

Rothstein, J.M. (1989). On defining subjective and objective measurements. Physical Therapy, 69, 577-579.

Rothstein, J.M., Campbell, S.K., Echternach, J.L., Jette, A.M., Knecht, H.G., and Rose, S.J. (1991). Standards for tests and measurements in physical therapy practice, Physical Therapy, 71, 589-621.

Rothstein,J.M. and Echternach, J.L. (1993). Primer on measurement: An introductory guide to measurement issues, American Physical Therapy Association, Alexandria, Va.

Rothstein, J.M. (1994). Disability and our identity (editorial). Physical Therapy, 74, 375-378.

Seto, J.L., Orofino, A.S., Morrissey, M.C., Medieros, J.M., and Mason, W.J. (1988). Assessment of quadriceps/ hamstring strength, knee ligament stability, functional and sports activity levels five years after anterior cruciate ligament reconstruction. American Journal of Sports Medicine, 16, 170-180.

Shelbourne, K.D., Rettig, A.C., McCarroll, J.R., Vogel, A., Kuhn, D., and Bisesi, M.A. (1987). Functional ability in athletes with an anterior cruciate ligament deficiency (abstract). American Journal of Sports Medicine, 15, 628.

Shrout, P.E. and Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86, 420-428.

Smyth, C.J., Velayos, E.E., and Hlad, Jr., C.J. (1963). A method for measuring swelling of the hands and feet. Part I: normal variations and applications in inflammatory joint diseases. Acta Rheumatologica Scandinavica, 9, 293-305.

Specht, D.A. (1975). On the evaluation of causal models. Social Science Research, 4, 113-133.

Tegner, Y. and Lysholm, J. (1985). Rating systems in the evaluation of knee ligament injuries. Clinical Orthopaedics & Related Research, 198, 43-9.

Tegner, Y., Lysholm, J., Lysholm, M., and Gillquist, J. (1986). A performance test to monitor rehabilitation and evaluate anterior cruciate injuries. American Journal of Sports Medicine, 14, 156-159.

Tegner, Y., Lysholm, J., Odensten, M., and Gillquist, J. (1988). Evaluation of cruciate ligament injuries. A review. Acta Orthopaedica Scandinavica, 59, 336-341.

Tibone, J.E., Antich, T.J., Fanton, G.S., Moynes, D.M., Perry, J. (1986). Functional analysis of anterior

cruciate ligament instability. American Journal of Sports Medicine, 14, 276-283.

Tibone, J.E., and Antich, T.J. (1988): A biomechanical analysis of anterior cruciate ligament reconstruction with the patellar tendon. American Journal of Sports Medicine, 16, 332-335.

Walla, D.J., Albright, J.P., McAuley, E., Martin, R.K., Eldridge, V., and El-Khoury, G. (1985). Hamstring control and the unstable anterior cruciate deficient knee. American Journal of Sports Medicine, 13, 34-39.

Wilkinson, G.B. and Horn-Kingery, H.M. (1993). Treatment of the inversion ankle sprain: Comparison of different modes of compression and cryotherapy. Journal of Orthopaedic and Sports Physical Therapy, 17, 240-246.

Wilks, K.E., Romaniello, W.T., Soscia, S.M., Arrigo, C.A., and Andrews, J.R. (1994). The relationship between subjective knee scores, isokinetic testing, and functional testing in the ACL-reconstructed knee. Journal of Orthopaedic and Sports Physical Therapy, 20, 60-72.

World Health Organization (1980). International Classification of Impairments, Disabilities and Handicaps. Geneva.

Wright, S. (1934). The method of path coefficients. The Annals of Mathematical Statistics, 5, 161-215.

Wright, S (1960). Path coefficients and path regresions: Alternative or complementary concepts? Biometrics, 16, 189-202.

Appendix A

CONSENT TO PARTICIPATE IN A STUDY

Title: Validity of physical, perceptual, and performance measures following ankle inversion injury

We invite you to participate in a study of physical, performance, and perceptual measures used during rehabilitation following ankle inversion injury. We hope to learn whether athletes' responses to these activities provide useful information regarding their progress in rehabilitation and readiness to return to athletic participation. You were selected because you are an athlete who is currently undergoing rehabilitation for an ankle inversion injury.

Investigational Procedures

If you choose to participate in this study, you will be measured in conjunction with your regular rehabilitation sessions at the University of Virginia Sports Medicine Clinic. Rehabilitation sessions are normally conducted daily until you are ready to return to practice and competition, normally no longer than four to six weeks from the date of your injury. Measurements sessions will be conducted on several occasions in conjunction with your rehabilitation. These sessions will include range of motion and volume measurements. Following the measurement sessions, you will be asked to perform a number of motor activities (standing balance, shuttle run, figure 8 run, single leg hop, triple hop, stairs hop). These activities are commonly employed during rehabilitation to improve balance and coordination following lower extremity joint injuries among athletes. You will be allowed to select the tasks you will be measured on. If you find a task too difficult or uncomfortable, we ask that you stop immediately and tell the observer. Your performance on that task will not be measured that day. Following administration of the performance tasks, you will complete a short questionnaire addressing your impressions during testing.

Risks and Benefits

Research studies often involve some risks. The risks of this study are possible sprains, including re-injury, muscle strains, or bruises which could typically result from activities requiring running and jumping. In addition, it is possible in any situation that accidents or other harmful effects which are unknown may occur. Of course, we will take precautions to watch for and identify any harmful side effects.

If you participate in this study, you will receive frequent, personally supervised physical rehabilitation. At the conclusion of the study, you will also receive the results of your performance on the various tests. This information may help you determine whether you are ready to return to practice and competition. Also, your fellow athletes and the sports medicine community may benefit from knowing which measures are the most accurate indicators of functional ability following ankle inversion sprains.

Alternatives to Participating in This Study

The alternative is to not participate in this study.

If you decide to participate, you may withdraw from the study at any time by telling the investigator and leaving the room.

Neither your decision not to participate in this study nor your decision to stop and withdraw from a study once you decide to participate, will affect your grading, academic, or athletic standing at this university. Of course, we will tell you anything we learn during the study that may help you decide whether to continue participating.

Privacy of Records

Any information that we learn about you that can be individually traced to you will be used responsibly and will be protected against release to unauthorized people. In addition to the members of the health care staff who usually have access to your file, your records are likely to be shown to members of the US Food and Drug Administration.

If you sign this form, you have given us permission to release information to these other people. The results of this study may be published in the medical literature, but no publication will contain information which will identify you.

In the event anyone involved in this study is exposed to your blood or body fluids, your blood may be tested for evidence of hepatitis, AIDS, or other infections without your consent.

Payment

In the event you suffer physical injury directly resulting from the research procedures, no financial compensation for such things as lost wages, disability, or

discomfort is available, but medical treatment that is not covered by your insurance will be provided free of charge by the University of Virginia. If you have any questions concerning financial compensation for injuries caused by the experiment, you should talk to Rick Wilson at (804) 982-5450.

Conclusion

You are making a decision whether you will participate in this study. If you sign this form, you are agreeing to participate based on your reading and understanding this form. If you have any questions, please ask Rick Wilson or Joe Gieck at (804) 982-5450.

If you have any questions regarding research subjects' rights, please contact Dr. Jerry Short, Chair of the University Committee for the Protection of Human Subjects, (804) 924-7471.

You will receive an unsigned copy of this form to keep.

Witness	Subject
Member of Research Team	Date

APPENDIX B

Ankle Study: Data Collection Form

DOI________ Subject _______

Sport: Date _______

Physical Measures

Volume ______/______ ______/______ (inj / nl)

ROM ______/______ ______/______ ______/______ (inj / nl)

Functional Activities

Ambulation			
Non WB	no	yes	
Partial WB	no	yes	Date:
Full WB	no	yes	Date:
40 m. Run	no	yes	
Figure 8 Run	no	yes	
Single hop	no	yes	
Triple hop	no	yes	
Stairs Hop	no	yes	
Return to:			
Limited participation	no	yes	Date:
Full Participation	no	yes	Date:

APPENDIX C

Ankle Study: Functional Ability Rating Sheet

Subject________   Date_________

You have just completed a test designed to assess your functional ability. Please respond to the following question by making a single vertical mark on the line below.

Compared to your usual level of performance, how would you rate your athletic ability today?

extremely limited ____ no limitations ; normal ability

Appendix D

Ankle Study: Performance Rating Sheet

Test_______ **Subject _______**

Condition_______ **Date _______**

You have just completed a test designed to measure your physical performance. Please make a single vertical mark on the line to indicate your impressions during the test.

1. How much your ankle injury interfere with your performance during this test?

None ____ Maximum

2. The amount of apprehension or "guarding" I experienced during this test was:

None ____ Maximum

3. The amount of effort I put forth on this test was:

None ____ Maximum

4. The amount of ankle pain or discomfort I experienced during this test was:

None ____ Maximum

5. The amount of ankle instability or "weakness" I experienced during this test was:

None ____ Maximum

Appendix E: Repeated Measures Analysis of Variance and Intraclass Correlation Coefficients

Source of Variance	SS	DF	MS	F	p	$ICC_{2,1}$
Between Subjects	4509.89	18	250.55			
Within Subjects						
PAIN	6.74	1	6.74	.52	.48	0.90
Residual	231.26	18	12.85			
Between Subjects	7012.89	20	350.64			
Within Subjects						
ROM	6.89	2	3.44	.60	.55	0.95
Residual	228.44	40	5.71			
Between Subjects	8189.27	20	409.46			
Within Subjects						
ROMLOSS	30.13	2	15.06	.83	.46	0.88
Residual	728.54	40	18.21			
Between Subjects	2262972.62	20	113148.63			
Within Subjects						
VOLUME	314.88	1	314.88	4.76	.04	0.99
Residual	1322.62	20	66.13			
Between Subjects	106033.33	20	5301.67			
Within Subjects						
SWELLING	5.36	1	5.36	.04	.85	0.95
Residual	2807.14	20	140.36			
Between Subjects	14805.06	16	925.32			
Within Subjects						
ATHABILITY	62.24	1	62.24	.87	.36	0.86
Residual	1141.76	16	71.36			
Between Subjects						
Within Subjects						
ACTIVITY	No Variance					
Residual						

ATHAB1A	ATHAB1B	INJVOL1A	INJVOL1B	NLVOL1A	NLVOL1B	INROM1A	INROM1B	INROM1C	NLROM1	NLROM1	NLROM1
0	0	1310	1300	1225	1215	30	28	32	75	75	79
34	37	1200	1185	1160	1170	42	42	43	70	65	66
83		1830	1820	1810	1830	68	71	68	75	75	77
22	44	1420	1430	1305	1310	51	54	50	47	42	60
75	76	1350	1355	1330	1360	64	65	65	58	61	62
0		1140	1140	1070	1070	55	49	51	69	72	70
66	67	1440	1440	1380	1370	62	59	62	70	68	68
36		1675	1690	1550	1530	40	43	43	55	54	52
2	4	1895	1900	1700	1710	36	35	33	51	54	52
37	30	1330	1345	1340	1325	60	56	62	64	65	61
73	65	1705	1740	1650	1680	55	61	58	60	63	64
58	23	1525	1540	1430	1450	64	66	65	70	65	65
65	59	1485	1495	1390	1405	40	46	39	40	41	44
63	54	1280	1290	1280	1280	45	44	41	58	57	60
82	83	1260	1270	1225	1220	58	58	60	63	60	64
71	75	1185	1190	1140	1140	42	48	52	45	53	50
62	63	960	970	915	920	46	47	49	61	61	63
32	43	1310	1315	1280	1300	41	41	37	42	39	45
30	15	1475	1490	1360	1370	52	50	50	43	49	53
64	61	1150	1140	1140	1140	44	48	48	61	61	67
49	40	1255	1250	1145	1160	64	61	67	68	66	62

PAIN1A	PAIN1B	RETURN	MOTOR2	ATHAB2A	ATHAB2B	INJVOL2A	INJVOL2B	INROM2A	INROM2B	INROM2C	SWELL1
0	0	24	1	1		1250		47	49	54	85
4	2	10	7	94		1170		65	61	60	27.5
21	6	4									5
		6									117.5
19	23	5									7.5
44	39	23	4	56	63	1090		65	64	64	70
6	5	3	8	96		1400	1395	52	52	54	65
8	18	17	8	78	68	1595	1595	43	46	49	142.5
		19	2	17		1725	1750	40	41	45	192.5
23	25	17	6	56		1330	1340	49	50	48	5
0	0	11	8	93	73	1715	1720	50	48	51	57.5
15	15	16	7	74	94	1500	1510	70	66	73	92.5
0	0	10									92.5
25	28	10	7	86	90	1240	1260	48	53	53	5
1	1	4	8	94	93	1220	1235	54	55	55	42.5
2	2	7	8	99	97	1170	1180	59	59	59	47.5
17	12	12									47.5
5	6	17	8	94	92	1290	1300	43	44	45	22.5
19	11	21									117.5
6	7	8									5
3	2	5									100

ROMLOS	MOTOR1	ROMLOS	SWELL2	DOB	DOI	OCCAS1	OCCAS2	AGE	HRSPOS	INTERVAL	
-46.3333	0	-26.3333	30	5/19/76	9/26/94	9/29/94	10/4/94	18.36904	64.41667	5.305556	
-24.6667	2	-5	5	5/24/74	9/29/94	10/1/94	10/5/94	20.36627	40.66667	4.284722	
-6.66667	8			5/2/73	10/1/94	10/4/94		21.43196	70.25		
2	7										
4.333333	7			1/11/72	10/10/94	10/13/94		22.76364	71.66667		
-18.6667	1	-6	20	5/6/76	10/11/94	10/13/94	10/20/94	18.44578	48.33333	6.944444	
-7.66667	7	-16	22.5	10/19/71	10/17/94	10/19/94		23.01256	49.5		
-11.6667	6	-7.66667	55	12/17/71	10/17/94	10/20/94	10/27/94	22.85126	66.5	7.03125	
-17.6667	1	-10.3333	32.5		10/21/94	10/24/94	10/31/94		63	7.041667	
-4	3	-14.3333	2.5	5/28/73	10/17/94	10/21/94	10/28/94	21.40468	87	7.041667	
-4.33333	7	-12.6667	52.5	11/21/75	10/21/94	10/23/94	10/28/94	18.93057	45.25	5.125	
-1.66667	2	3	65	7/3/75	10/30/94	11/2/94	11/9/94	19.34087	76	6.958333	
0	5			3/14/75	10/30/94	11/2/94		19.64623	65.75		
-15	4	-7	30	5/30/76	11/5/94	11/9/94	11/17/94	18.44912	88.75	8.041667	
-3.66667	7	-7.66667	5	2/8/73	11/8/94	11/11/94	11/17/94	21.76358	71.5	6	
-2	7	9.666667	35	5/10/76	11/14/94	11/17/94	11/22/94	18.52765	74.25	5	
-14.3333	4			6/11/73	11/21/94	11/25/94		21.46264	90.75		
-2.33333	5	2	5	7/13/75	12/1/94	12/5/94	12/13/94	19.40181	91.41667	8.21875	
2.333333	7			1/11/77	1/29/95	2/1/95		18.06221	71.5		
-16.3333	6			12/27/73	2/3/95	2/5/95		21.11912	48.75		
-1.33333	7				3/1/95	3/4/95			71.75		